

Analisis Sentimen Terhadap KPU 2024 Berdasarkan Tweet Media Sosial Twitter Menggunakan Algoritma Naïve Bayes

Dion Parisda Ray, Firman Noor Hasan, Ahmad Rizal Dzikrillah*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹dion7472@gmail.com, ²firman.noorhasan@uhamka.ac.id, ^{3,*}ahmadrizaldzikrillah@gmail.com

Email Penulis Korespondensi: ahmadrizaldzikrillah@gmail.com

Abstrak—Perkembangan teknologi saat ini sangat pesat membuat penyebaran informasi semakin cepat, penyebaran informasi sangat mudah didapat di media sosial seperti Twitter. Media sosial Twitter sendiri menyediakan fitur bagi penggunanya untuk dapat mengirim dan membaca informasi baik berupa teks maupun video. Masyarakat cemas bahwa pemilu 2024 dapat menciptakan ketegangan dan disintegrasi jika hasilnya tidak sesuai dengan pilihan suara mereka. Pemilu merupakan momen yang sangat penting bagi bangsa Indonesia dalam memilih pemimpin, dalam hal ini “KPU 2024” sebagai penyelenggara diharapkan mampu menyelenggarakan pemilu agar berjalan dengan baik. Data Twitter yang dikumpulkan dengan kata kunci “KPU 2024” diperoleh total 3057 dataset, dilanjutkan dengan proses pembersihan yang menghasilkan 715 dataset. Tujuan dari penelitian ini adalah untuk mengetahui berapa jumlah komentar tweet positif dan negatif serta untuk menunjukkan keakuratan penerapan metode Naïve Bayes. Hasil akurasi yang diberikan oleh algoritma Naïve Bayes sebesar 67,13% dengan presisi sebesar 66,04% dan recall sebesar 100,00%. Penelitian ini dilakukan untuk melihat sentimen masyarakat terhadap “KPU 2024” nanti. Hasil evaluasi pada matriks konfusi diperoleh true positif sebesar 457 dan true negative sebesar 235.

Kata Kunci: Analisis Sentimen; KPU 2024; Twitter; Rapidminer; Naïve Bayes

Abstract—The development of technology is currently very rapid making the dissemination of information faster, the dissemination of information is very easy to get on social media such as Twitter. Twitter social media itself provides features for its users to be able to send and read information in the form of text or video. Elections are a very important moment for the Indonesian people in choosing leaders, in this case the “2024 KPU” as the organizer is expected to be able to run the elections so that they run well. Twitter data collected with the keyword “KPU 2024” obtained a total of 3057 datasets, followed by a cleansing process which produced 715 datasets. The aim of this research is to find out how many positive and negative tweets comments and to indicate the accuracy of the implementation of the Naïve Bayes method. The accuracy results given by the Naïve Bayes algorithm are 67.13% with a precision of 66.04% and a recall of 100.00%. This research was conducted to see public sentiment towards the “2024 KPU” later. Evaluation results in the confusion matrix obtained true positives of 457 and true negatives of 235.

Keywords: Setiment Analysis; KPU 2024; Twitter; Rapidminer; Naïve Bayes

1. PENDAHULUAN

Twitter adalah salah satu media sosial penghubung yang digemari oleh semua warga di global. Hal ini terlihat asal meningkatnya total pemakai *twitter* yang tercatat pada seluruh global, termasuk Indonesia. Waktu ini *twitter* mempunyai total pemakai aktif sebesar 313 juta per bulan. di tahun 2016 pemakai *twitter* akan memberikan berita atau ulasan terkini tentang perihal yang sedang sebagai topik utama di global. Hal-hal yang ketika ini menjadi topik utama dan poly serta tak jarang ulasan oleh pemakai akan menyebabkan suatu permasalahan atau trending topik di media sosial khususnya *twitter*. [1]. Warga semakin cakap menggunakan media sosial dengan banyak sekali cara, mulai berasal mengakses isu, komunikasi, belajar, berbelanja serta lain sebagainya. Media sosial hadir untuk kemudahan serta perubahan bagi warga, sebagai akibatnya pengguna media sosial bertambah banyak [2]. Hal ini bisa terjadi sebab berkembang serta canggihnya pertumbuhan teknologi pada global waktu ini, misalnya masyarakat bisa mengakses berita tentang apa yang terjadi di luar negeri secara langsung lewat media sosial tentang keadaan disana. Menggunakan pertumbuhan teknologi di bidang internet, penyebaran isu dan komunikasi semakin tinggi secara terus-menerus. Media sosial yang mensupport kemudahan peluasan isu ini [3]. Termasuk media sosial terpopuler dipergunakan oleh warga bangsa merupakan *twitter*. Menjadi perbedaan, ibarat media sosial Instagram serta Facebook bukan diteliti pada pengkaji ini sebab Instagram berpusat pada foto serta pemakai Facebook tidak memposting yang berpusat pada teks. [4]. Maka berasal itu penelitian ini memakai *twitter* menjadi bahan dalam pengolahan dan pemrosesan data nantinya.

Di bangsa ini sampai 130 juta pemakai dengan penetrasi sampai 49%. Penelitian asal Reportal Data menunjukkan bahwa banyak pengguna sosial media pada tanah air mencapai 191,4 juta per Januari 2022. Termasuk perangkat sosial *Twitter* KPU yang merupakan cara pemutakhiran aktivitas perihal ulasan pemakai *twitter* di Indonesia. Pengkajian ini dipusatkan pada perangkat sosial *Twitter* buat mengkaji ulasan-ulasan perihal KPU [5]. Pemilu yang diselenggarakan secara rutin adalah suatu parameter kerakyatan pada bangsa. semenjak perbaikan, bangsa Indonesia sudah menyelenggarakan pemilu sebesar 5 kali, yakni pemilu 1999, 2004, 2009, 2014, serta 2019. [6]. Pemilu adalah sarana buat mewujudkan kedaulatan masyarakat. Pemilu juga ialah ajang persaingan. Menang atau kalahnya seorang calon dipengaruhi sang masyarakat melalui prosedur pengambilan suara. memilih opsi ketika pemilu adalah hak setiap warga negara. menjadi perangkat yang amat istimewa dalam rangka memilih serta berpartisipasi dalam memilih wakil serta pemimpin rakyat yang bakal duduk pada pemerintahan, warga negara diberi kesempatan dalam pemilu buat menentukan pejabat pemerintah yang betul-betul diklaim bisa menyuarakan kehendaknya. [7]. Komisi Pemilihan umum atau KPU merupakan forum pengurus pemilu yang berjiwa nasional, permanen, serta berdikari pada penyelenggara pemilu. daerah kerja KPU mencakup semua daerah bangsa indonesia. KPU melaksanakan perannya secara berkelanjutan serta dengan

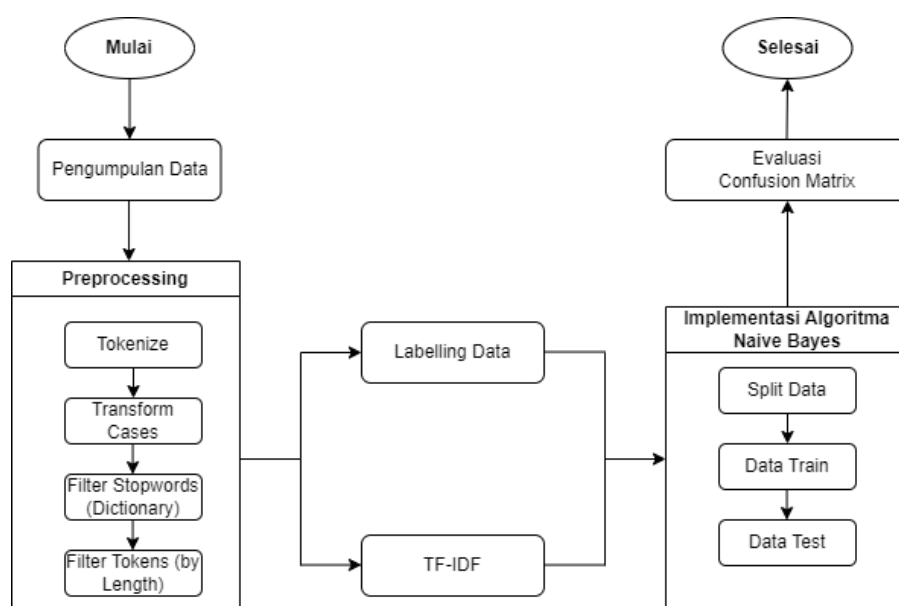
menyelenggarakan pemilu, KPU independen asal dampak golongan manapun yang terkait menggunakan aplikasi tugas serta tanggung jawab. [8].

Analisis sentimen merupakan proses menemukan kesamaan individu lewat kalimat ataupun teks atau penggalan data kata buat memperoleh isu perihal sentimen yang umumnya ditulis di sosial media, baik positif, negatif, atau netral. Tahapan analisis ini umumnya dilakukan secara otomatis melewati jumlah software atau platform yang mensupport analisis teks [9]. Informasi berbasis teks poly tersedia di internet dalam format forum media sosial, website personal, platform medsos serta website komentar barang. Dengan bantuan analisis sentimen, isu yang tadinya tidak tersusun dapat diubah menjadi data yang lebih tersusun [10]. Sebelumnya sudah ada beberapa penelitian terkait sentimen analisis menggunakan memakai prosedur pemecahan *Naive Bayes*. prosedur pemecahan *Naive Bayes* dipilih sebab metode pembagian terstruktur mengenai ini mempunyai formula yang sederhana serta simpel dipraktikkan juga mempunyai ketepatan yang relatif tinggi dibandingkan menggunakan prosedur lainnya. [11]. *Naive Bayes* dipercaya menjadi cara yang berpotensi bagus buat pemrosesan klasifikasi data tersusun daripada cara klasifikasi tersusun lainnya di dalam hal ketepatan dan komputasi. Penelitian yang dilakukan oleh Hasan serta Dwijayanti menghasilkan akurasi dalam pengujian pemecahan pada metode *Naive Bayes Classifier* yaitu 92,5% [12]. Masyarakat mengalami kekhawatiran terkait KPU 2024 sebagai lembaga penyelenggara pemilu, merujuk pada kecurangan yang terjadi pada pemilu sebelumnya. Penelitian ini bertujuan untuk menggali sentimen masyarakat yang terungkap melalui tweet di media sosial Twitter, baik dalam bentuk positif maupun negatif. Sentimen yang tercermin dalam interaksi online ini dianggap sebagai indikator kepercayaan masyarakat terhadap KPU 2024 sebagai penyelenggara pemilu. Peneliti yang peneliti lakukan dengan peneliti sebelumnya terdapat beberapa perbedaan. Mengenai apa yang melainkan penelitian ini dengan penelitian terdahulu berbeda ialah topik serta *Filter Tokens (by Length)* dalam *preprocessing* sedangkan penelitian sebelumnya menggunakan *Stemming*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian diawali menggunakan pengumpulan data yang diperoleh dari *crawling* di sosial media *twitter* dengan *rapidminer*. *RapidMiner* merupakan perangkat yang bersifat terbuka (open source). *RapidMiner* merupakan penyelesaian buat mengkaji data *mining*, *text mining* dan analisis prediktif, *RapidMiner* memanfaatkan beragam cara deskriptif serta prediktif buat membagikan pengetahuan oleh pemakai sehingga dapat mengambil keputusan terbaik [13]. *Naive Bayes* adalah pendekatan tanpa pedoman. Algoritma ini memanfaatkan bidang matematika yang diakui sebagai teori kemungkinan untuk menemukan potensi terbesar untuk penjabaran, menggunakan melihat frekuensi setiap penjabaran dalam data pelatihan. *Naive Bayes* merupakan suatu teknik yang diterapkan buat menentukan kelas berasal setiap soal, yang telah dibagi sesuai masing-masing soal. perhitungan numerik sesuai pendekatan kelompok [15]. Berikut alur penelitian pada Gambar 1 dibawah.



Gambar 1. Alur Penelitian

Penelitian ini mengawali proses pengumpulan data dengan memanfaatkan tools *RapidMiner* yang telah dilengkapi dengan API untuk mengekstraksi data dari platform *Twitter*. Peneliti menggunakan aplikasi *RapidMiner* dengan API yang telah tersedia untuk mengakses dan mengambil data dari media sosial *Twitter*. Setelah mendapatkan data, tahap pra-pemrosesan dilakukan untuk membersihkan dan mempersiapkan data yang telah terkumpul. Proses ini, umumnya disebut *preprocessing*, memiliki beberapa langkah, seperti *cleansing* untuk membersihkan teks tweet dari elemen-elemen seperti

link dan istilah RT (Re-tweet). Selanjutnya, tahapan tokenize digunakan untuk memisahkan kata-kata menjadi token yang memiliki makna. Transform Cases berfungsi untuk mengubah format huruf, seperti mengubah ke kapital atau huruf kecil. Stopword digunakan untuk menghapus kata-kata yang dianggap tidak penting. Filter Tokens bertujuan untuk menghilangkan kata-kata dengan jumlah huruf yang terlalu sedikit atau terlalu banyak. Setelah proses pra-pemrosesan selesai, data dibelabeli secara manual oleh peneliti sebagai tahap pebelaban data. Selanjutnya, diterapkan teknik TF-IDF (Term Frequency Inverse Document Frequency) untuk mengukur seberapa sering sebuah kata muncul dalam suatu dokumen. Implementasi algoritma Naïve Bayes dilakukan dengan membagi data yang sudah dilabeli menjadi dua bagian, yakni data latih dan data uji, menggunakan operator split data. Terakhir, evaluasi dilakukan dengan menerapkan confusion matrix pada hasil implementasi algoritma Naïve Bayes. Hasil dari evaluasi ini memberikan nilai-nilai seperti Precision, Recall, dan Accuracy untuk mengukur kinerja algoritma dalam melakukan klasifikasi sentimen.

3. HASIL DAN PEMBAHASAN

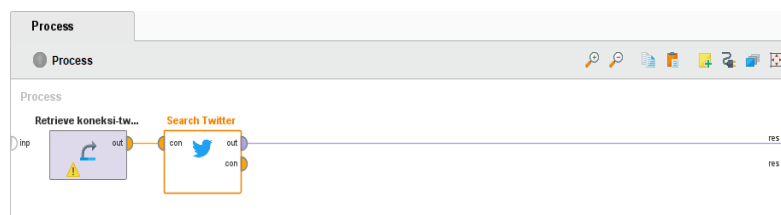
3.1 Pengumpulan Data

Peneliti mengumpulkan data di sosial media *twitter* terhadap *tweet* yang berisi *tweet* masyarakat menggunakan istilah kunci pencarian “KPU 2024” menggunakan aplikasi *RapidMiner* dengan *API* yang sudah tersedia. Berikut ilustrasi pengambilan data pada Gambar 2.



Gambar 2. Ilustrasi Pengambilan Data

Pada Gambar 2 diatas adalah ilustrasi pengambilan data yang didapat melalui media sosial *twitter*, peneliti melakukan pengambilan data selama 3 bulan yaitu pada Agustus, September, dan November 2023 selanjutnya diproses menggunakan *tools RapidMiner* dan disimpan dalam format file CSV dengan nama file data_kpu.csv. Berikut tahapan crawling pada Gambar 3 dibawah.



Gambar 3. Crawling Data

Tahapan proses penelitian yang dilakukan, dimulai dari menggunakan dua buah operator yaitu *retrieve API* yang terhubung dengan *Search Twitter* untuk *crawling* data di *tools RapidMiner*. *Crawling* data yang dilaksanakan yakni melalui memilih kata kunci yang diperlukan buat pemungutan data, pada hal ini istilah kunci ditulis oleh peneliti yakni “KPU 2024” menggunakan operator *Search Twitter* yang sudah terkoneksi dengan *API RapidMiner* yaitu operator *retrieve*. Setelah itu, peneliti mendapatkan sebanyak 3057 data lalu disimpan di dalam file csv buat dilakukan cara selanjutnya. Berikutnya hasil penarikan data pada Gambar 4 dibawah.

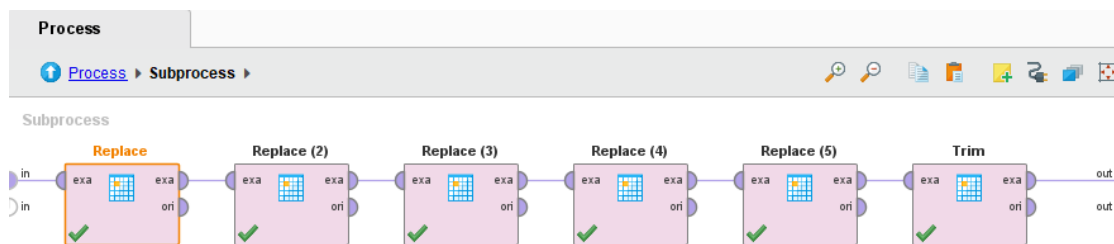
	A	B
1	Bila MK ngotot akan tetap memutuskan mengubah Pemilu jadi sistem proporsional tertutup, maka itu juga ber	1,66E+18
2	Saya yakin, dlm menyusun DCS, Parpol & Caleg berasumsi sistem pemilu tidak diubah, tetap sistem terbuka. Ki	1,66E+18
3	Bentuk Kotak Suara Pemilu 2024, KPU: Tetap Pakai Karton Tapi Lebih Kuat https://t.co/7ZOzCevbw9	1,66E+18
4	Pahami Aturan Pemilu 2024, Bawaslu tegaskan kepada para peserta Pemilu untuk pahami dan laksanakan atur	1,66E+18
5	Pahami Aturan Pemilu, Bawaslu tegaskan kepada para peserta Pemilu untuk pahami dan laksanakan aturan Pe	1,66E+18
6	Istana ungkap Presiden Jokowi tak mau ada polarisasi di pemilu 2024Pancasila Flamengo Jokic#PemiluSerenta	1,66E+18
7	RT @ForRezim: #SaatnyaBukaTopengJokowi Mestinya Jokui Cawe cawe agar?TNI_Polri_Harus_Netral cara cu	1,66E+18
8	Pj Walikota Pekanbaru dorong masyarakat berikan hak suara dalam pemilu 2024Pancasila Flamengo#PemiluSe	1,66E+18
9	RT @KPU_ID: Siber Bareskrim Polri, BRIN, Kominfo, BIN RI, dan BSSN. Betty berharap KPU mempunyai Satu Pe	1,66E+18
10	RT @KPU_ID: Selain itu ada beberapa sistem informasi dalam tahapan pelaksanaan Pemilu serentak 2024, ya	1,66E+18
3057	RT @ForRezim: Awas! rezim ulangi curang pilpres 2024? mrk merasa tdk akan ada yg berani melawan! Awas K	1,66E+18

Gambar 4. Tampilan file csv

Pada tampilan file csv diatas merupakan hasil penarikan data menggunakan *tools RapidMiner* dengan *API* yang sudah tersedia.

3.2 Cleansing

Pada tahap ini, dilakukan penghapusan karakter atau elemen seperti simbol-simbol [!~@#&*%();",.-]. sehabis proses pencucian, dataset akan tersisa sinkron menggunakan kebutuhan diinginkan.



Gambar 5. Tahapan pembersihan memakai *replace* serta *trim*

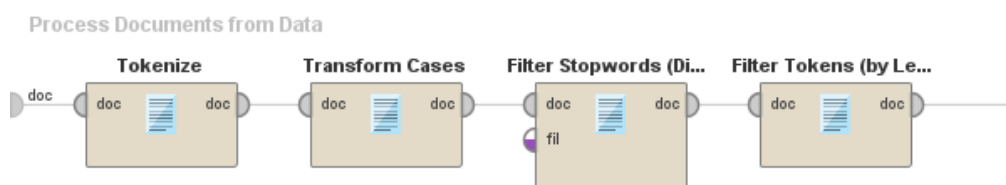
Pada Gambar 5 diatas merupakan tahapan *cleansing* dimana menggunakan lima operator *replace* serta *trim*. Pertama, penggunaan operator *replace* dimaksudkan untuk mengeliminasi kata "RT" dalam sebuah kalimat. Kedua, fungsi dari operator *replace 2* adalah menghapus URL. Selanjutnya, operator *replace 3* dapat diterapkan untuk menghapus segala kemunculan @username dalam suatu teks. Operator *replace 4* berperan dalam menghapus tanda pagar (#) atau hashtag. Selain itu, operator *replace 5* dapat digunakan untuk menghilangkan simbol atau karakter yang tidak relevan dalam suatu teks. Terakhir, operator *trim* berguna untuk menghapus baris kosong dalam teks tersebut. Dengan mengkombinasikan keenam operator ini, manipulasi teks dapat dilakukan sesuai dengan kebutuhan yang diinginkan.

Setelah melakukan *cleansing*, berikut merupakan yang akan terjadi proses pencucian tweet yang dilakukan pada *RapidMiner* terdapat pada Table 1 dibawah.

Table 1. *Cleansing* Dataset diatas

Sebelum	Sesudah
RT @Malika6027: VIRAL Seorang Bapak usulkan kepada KPU agar setiap Capres yang ikut dlm kontestasi Pemilu nanti DI TES MEMBACA ALQUR'AN	VIRAL Seorang Bapak usulkan kepada KPU agar setiap Capres yang ikut dlm kontestasi Pemilu nanti DI TES MEMBACA ALQUR'AN
RT @tvOneNews: KPU Sedang Nanti Vonis majelis hukum perkara Sangkaan Bacaleg memakai Dana Narkoba buat Pemilu https://t.co/N4SyznV2NU	KPU Sedang Nanti Vonis majelis hukum perkara Sangkaan Bacaleg memakai Dana Narkoba buat Pemilu
Gawat ICW Tuding KPU Seludupkan Pasal yang Permudah Narapidana Korupsi Ikut Pemilu https://t.co/RB4YangejKx4	Gawat ICW Tuding KPU Seludupkan Pasal yang Permudah Narapidana Korupsi Ikut Pemilu
RT @KompasTV: Dugaan Aliran Dana Narkoba ke Papol di Pemilu	Dugaan Aliran Dana Narkoba ke Papol di Pemilu

3.3 Preprocessing



Gambar 6. Tahapan *preprocessing*

Pada gambar 6 diatas merupakan tahapan *preprocessing*. Pada proses ini terdapat operator *tokenize*, *transform cases*, *filter stopword (dictionary)*, dan *filter tokens (by length)*. *Tokenize* sendiri berfungsi untuk memisahkan satu kata dengan kata lainnya, *transform cases* untuk mengubah *font*-nya misalkan mengubah ke kapital atau mengubah huruf kecil semua. *Stopword* untuk menghapus kata-kata yang tidak penting. Terakhir adalah *filter token (by length)* untuk menghapus kata-kata yang hanya mengandung sedikit huruf atau kata-kata yang mengandung terlalu banyak huruf.

3.3.1 Tokenize

Langkah pertama dari *preprocessing* adalah *tokenize*. *Tokenizing* merupakan cara analisis teks yang mulanya berbentuk kalimat-kalimat yang mengandung istilah-istilah [16]. Tujuannya adalah untuk mengelompokkan kemunculan sebuah kata pada proses selanjutnya dan sebuah kata pada proses selanjutnya [17].

Table 2. Hasil *tokenize*

Sebelum	Sesudah
---------	---------

VIRAL Seorang Bapak usulkan kepada KPU agar setiap Capres yang ikut dlm kontestasi Pemilu nanti DI TES MEMBACA ALQUR'AN	['VIRAL', 'Seorang', 'Bapak', 'usulkan', 'kepada', 'KPU', 'agar', 'setiap', 'Capres', 'yang', 'ikut', 'dlm', 'kontestasi', 'Pemilu', 'nanti', 'DI', 'TES', 'MEMBACA', 'ALQUR'AN']
KPU Sedang Nanti Vonis majelis hukum perkara Sangkaan Bacaleg memakai Dana Narkoba buat Pemilu	['KPU', 'Sedang', 'Nanti', 'Vonis', 'majelis', 'hukum', 'perkara', 'Sangkaan', 'Bacaleg', 'memakai', 'Dana', 'Narkoba', 'buat', 'Pemilu']
Gawat ICW Tuding KPU Seludupkan Pasal yang Permudah Narapidana Korupsi Ikut Pemilu	['Gawat', 'ICW', 'Tuding', 'KPU', 'Seludupkan', 'Pasal', 'yang', 'Permudah', 'Narapidana', 'Korupsi', 'Ikut', 'Pemilu']
Dugaan Aliran Dana Narkoba ke Parpol di Pemilu	['Dugaan', 'Aliran', 'Dana', 'Narkoba', 'ke', 'Parpol', 'di', 'Pemilu']

3.3.2 Transform Cases

Langkah kedua yaitu *transform cases*. *Transform Cases* berfungsi untuk mengubah *font*-nya misalkan mengubah ke kapital atau mengubah huruf kecil semua. Terlihat pada Table 3 kolom sebelah kiri merupakan kalimat yang memiliki beberapa huruf kapital dan kolom sebelah kanan merupakan kalimat kecil sama rata.

Table 3. Hasil *Transform Cases*

Sebelum	Sesudah
VIRAL Seorang Bapak usulkan kepada KPU agar setiap Capres yang ikut dlm kontestasi Pemilu nanti DI TES MEMBACA ALQUR'AN	viral seorang bapak usulkan kepada kpu agar setiap capres yang ikut dlm kontestasi pemilu nanti di tes membaca alqur'an
KPU Sedang Nanti Vonis majelis hukum perkara Sangkaan Bacaleg memakai Dana Narkoba buat Pemilu	kpu sedang nanti vonis majelis hukum perkara sangkaan bacaleg memakai dana narkoba buat pemilu
Gawat ICW Tuding KPU Seludupkan Pasal yang Permudah Narapidana Korupsi Ikut Pemilu	gawat icw tuding kpu seludupkan pasal yang permudah narapidana korupsi ikut pemilu
Dugaan Aliran Dana Narkoba ke Parpol di Pemilu	dugaan aliran dana narkoba ke parpol di pemilu

3.3.3 Filter Stopwords (Dictionary)

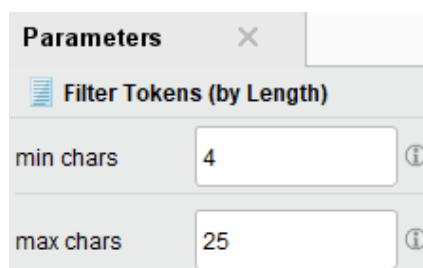
filter yang dipergunakan pada *RapidMiner* memakai bahasa Indonesia yang bisa didapatkan di website kaggle.com.

Table 4. Hasil *Filter Stopwords (Dictionary)*

Sebelum	Sesudah
viral seorang bapak usulkan kepada kpu agar setiap capres yang ikut dlm kontestasi pemilu nanti di tes membaca alqur'an	viral seorang bapak usulkan kpu capres dlm kontestasi pemilu di tes membaca alqur a
kpu sedang nanti vonis majelis hukum perkara sangkaan bacaleg memakai dana narkoba buat pemilu	kpu sedang nanti vonis majelis hukum perkara sangkaan bacaleg memakai dana narkoba pemilu
gawat icw tuding kpu seludupkan pasal yang permudah narapidana korupsi ikut pemilu	gawat icw tuding kpu selundupkan pasal permudah narapidana korupsi ikut pemilu
dugaan aliran dana narkoba ke parpol di pemilu	dugaan aliran dana narkoba parpol pemilu

3.3.4 Filter Tokens (by Length)

Tujuan dari tahap ini merupakan untuk menghapus istilah-istilah yang hanya mengandung sedikit huruf atau kata-kata yang mengandung terlalu banyak huruf.



Gambar 7. Parameter *Filter Tokens (by Length)*

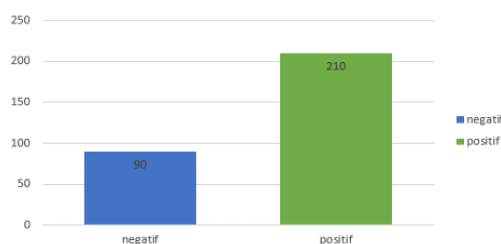
Pada cara ini istilah yang memiliki minim berasal 4 serta kian berasal 25 disingkirkan, sebagai akibatnya isu bisa lebih kentara sebab ketidakadaan istilah yang tidak informatif, berikut kata-kata hasil pemrosesan *Filter Tokens (by Length)*:

Table 5. Hasil Filter Tokens (by Length)

Sebelum	Sesudah
viral seorang bapak usulkan kepada kpu agar setiap capres yang ikut dlm kontestasi pemilu nanti di tes membaca alqur'an	viral seorang bapak usulkan kepada agar setiap capres yang ikut kontestasi pemilu nanti membaca alqur'an
kpu sedang nanti vonis majelis hukum perkara sangkaan bacaleg memakai dana narkoba buat pemilu	sedang nanti vonis majelis hukum perkara sangkaan bacaleg memakai dana narkoba buat pemilu
gawat icw tuding kpu seludupkan pasal yang permudah narapidana korupsi ikut pemilu	gawat tuding seludupkan pasal yang permudah narapidana korupsi ikut pemilu
dugaan aliran dana narkoba ke parpol di pemilu	dugaan aliran dana narkoba parpol pemilu

3.4 Pelabelan Data

Sesudah pembersihan ialah pelabelan data. Pelabelan data *twitter* pada tahapan ini dilakukan secara manual. Data yang sudah dikumpulkan lalu dibagi sebagai 2 kelas sentimen yaitu positif serta negatif [18], peneliti hanya melabeli 300 data dengan sentiment positif 210 dan negative 90, pertama yang dilakukan ialah secara manual data diberi label buat mengetahui mana yang negatif dan positif. Cara ini bermaksud buat training kepada prosedur pemecahan *naïve bayes* agar prosedur pemecahan tadi mampu melakukan pelabelan secara otomatis di tahap testing, berikut hasil pelabelan pada Gambar 8 dibawah.



Gambar 8. Visualisasi perbandingan positif dan negative

3.5 TF-IDF

Proses selanjutnya adalah pembobotan TF-IDF, proses ini dikerjakan buat menyampaikan kuantitas bobot pada istilah-istilah yang terdapat di *tweet* sesudah melalui termin pembersihan, buat mengerjakan pembobotan menggunakan TF-IDF ini memakai operator proses dokumen yang memuat operator berasal proses pembersihan yaitu, *tokenize*, *cases*, *stopword (dictionary)*, serta *filter tokens (by lengt)*. Berikut adalah gambar yang memberikan akibat berasal proses pembobotan TF-IDF yang menunjukkan total frekuensi kehadiran sebuah istilah di document. Berikut gambar dibawah mengenai hasil pembobotan TF-IDF.

Row No.	word	in documents	total ↓	in class (positif)	in class (negatif)
460	pemilu	218	227	169	58
106	dana	75	78	40	38
250	kampanye	52	54	36	18
412	narkoba	52	52	19	33
532	polri	48	48	27	21
46	bawaslu	42	42	34	8
91	cegah	32	32	30	2
15	aliran	30	30	9	21
529	politik	26	27	17	10
27	aturan	18	21	21	0
305	kota	10	18	18	0
61	berita	17	17	15	2
151	dukung	17	17	16	1
199	hoaks	16	16	15	1
441	partai	12	15	11	4

Gambar 9. Hasil proses pembobotan TF-IDF

Sesuai gambar diatas membagikan bahwa istilah yang paling poly ada yakni pemilu sebesar 218 istilah, dilanjutkan dengan istilah dana sebanyak 75 istilah, lalu istilah kampanye sebesar 52 istilah.

3.5.1 Visualisasi Wordcloud

Data ulasan yang sudah dilakukan lalu akan diimplementasi stopwords atau istilah awam yang tidak mempunyai arti seperti (yang, di, adalah). Sehingga, hanya bakal tertinggal istilah yang mempunyai arti seperti istilah kerja, istilah benda, serta lainnya. Sehabis kata bersih dari stopwords lalu bakal dilakukan visualisasi memakai *wordcloud* [19]. Sesudah

pembobotan TF-IDF peneliti pula membuat data visualization pada setiap istilah yang telah dilakukan pembobotan ke pada format *wordcloud* buat memperlihatkan 20 istilah yang mempunyai frekuensi kehadiran seringkali dipergunakan mirip Pada Gambar 10 dibawah.

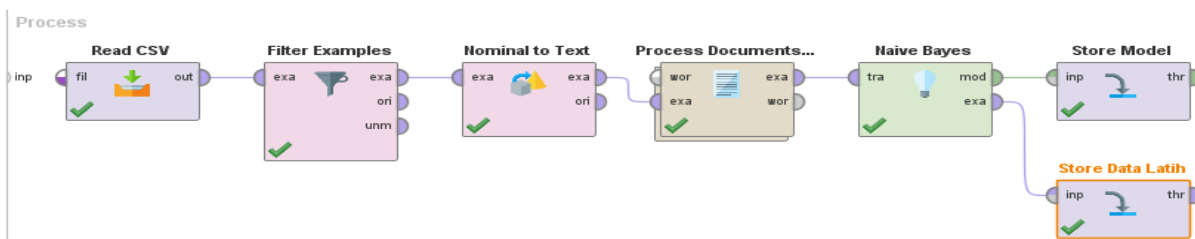


Gambar 10. Data visualization wordcloud

Sesuai gambar akibat data visualisation pada bentuk *wordcloud* diatas membagikan bahwa semakin banyak suatu ukuran istilah maka istilah tadi adalah total istilah yang teramat poly timbul. istilah yang seringkali timbul di dataset yaitu istilah pemilu, disusul menggunakan istilah dana, kampanye, narkoba, polri, bawaslu, aliran, politik, cegah, aturan kemudian kata hoaks. Dari data diatas dapat dilihat bahwa pemilu merupakan hal yang sangat penting untuk menghasilkan suara rakyat dalam memilih pemimpin.

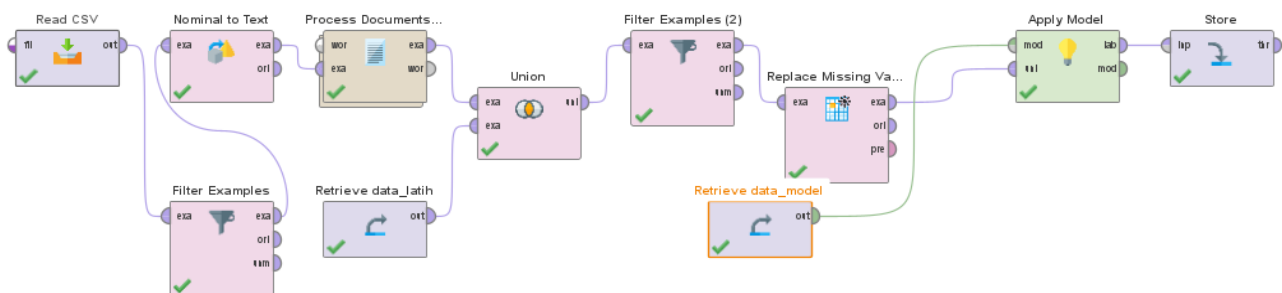
3.6 Implementasi Algoritma

Data yang telah melewati tahapan pemrosesan teks data kemudian akan melalui tahap klasifikasi memakai *Naïve Bayes Classifier* buat mengetahui data tersebut apakah bersifat positif atau negatif. Metode *Naïve Bayes Classifier* digunakan untuk kategorisasi, yaitu untuk mengetahui apakah komentar seseorang terkait suatu masalah atau objek cenderung berkategori negatif atau positif. [20]. Termin selanjutnya merupakan implementasi prosedur pemecahan naïve bayes. Cara memberi pembinaan data serta testing data peneliti memakai rasio 80% asal data yang akan menjadi training dan 20% asal data testing [21]. Berikut Pada Gambar 11 dibawah pemrosesannya.



Gambar 11. Proses pembuatan model Naïve Bayes dan data latih

Setelah mempersiapkan model Naïve Bayes dan data latih, langkah selanjutnya adalah melakukan klasifikasi menggunakan algoritma Naïve Bayes pada data uji yang berisi sentimen Twitter. Pertama, data uji difilter untuk mengambil hanya data yang belum memiliki label sentimen. Selanjutnya, dilakukan proses preprocessing menggunakan operator process document. Hasil dari operasi ini kemudian digabungkan dengan data latih menggunakan operator union. Tujuan dari penggabungan ini adalah untuk menyesuaikan atribut yang ada di antara kedua set data, mengingat kemungkinan perbedaan kata-kata di antara keduanya. Setelah digabungkan, data tersebut kembali difilter untuk memilih hanya data yang sentimennya masih kosong. Setelah persiapan data uji selesai, langkah berikutnya adalah melakukan proses klasifikasi. Model Naïve Bayes diaplikasikan pada data uji menggunakan operator apply, sehingga dilakukan klasifikasi sentimen secara otomatis. Setelah proses klasifikasi selesai, hasilnya berupa data yang sudah dilabeli sentimennya secara otomatis menggunakan algoritma Naïve Bayes. Berikut pada Gambar 12 dibawah proses pemberian label secara otomatis dengan menggunakan metode algoritma Naïve Bayes.



Gambar 12. Proses pelabelan secara otomatis

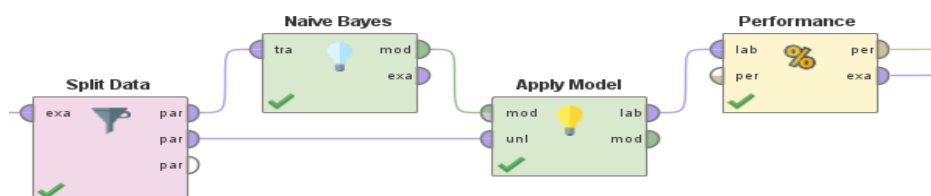
Setelah melalui proses otomatisasi pemberian label dengan menggunakan metode algoritma Naïve Bayes, hasilnya berupa prediksi sentimen dari algoritma tersebut. Pada tahap pemberian label otomatis oleh algoritma Naïve Bayes, algoritma tersebut telah menjalani proses pelatihan menggunakan data latih yang telah disiapkan sebelumnya oleh peneliti. Data latih ini awalnya telah diberi label secara manual. Setelah melalui fase pelatihan, algoritma Naïve Bayes dapat memahami pola yang telah dipelajari dari data latih tersebut. Dengan pemahaman tersebut, algoritma Naïve Bayes mampu memprediksi sentimen data baru ke dalam dua kelas, yaitu positif dan negatif. Berikut pada Gambar 13 dibawah hasil dari pemberian label secara otomatis dengan menggunakan metode algoritma Naïve Bayes.

text	prediction(Sentimen)
persiapan daftar pemilih pemilu rabu kamis tanggal aula lantai kantor provinsi jawa veteran semar...	positif
generasi muda wajib bantu sukseskan pemilu	positif
tana toraja anggarkan juta sewa gedung penyimpanan logistik pemilu	positif
download aplikasi tribunnews update berita terbaru	negatif
aliansi sekalbar deklarasikan dukung pemilu damai tolak politik identitas	positif
generasi muda papua diimbau jaga kamtibmas dukung pemilu damai	positif
anggarkan miliar pilgub	negatif
pbnu tolak politik identitas menjelang pemilu	positif
ulangkah pilpres	negatif

Gambar 13. Hasil dari proses pemberian label secara otomatis

3.7 Evaluasi

Langkah berikutnya setelah melakukan klasifikasi dengan algoritma Naïve Bayes adalah melaksanakan proses pengujian dan evaluasi performa algoritma tersebut. Dalam pengujian evaluasi ini, confusion matrix digunakan sebagai alat untuk mengevaluasi kinerja algoritma Naïve Bayes. Berikut pada Gambar 14 dibawah alur pengujian algoritma Naïve Bayes dengan confusion matrix.



Gambar 14. Alur pengujian algoritma naïve bayes

Dalam operator split data, peneliti akan memanfaatkan metode split data dengan perbandingan 80:20. Setelah melalui proses pengujian algoritma Naïve Bayes, hasilnya akan mencakup nilai-nilai Precision, Recall, dan Accuracy.

accuracy: 67.13%

	true positif	true negatif	class precision
pred. positif	457	235	66.04%
pred. negatif	0	23	100.00%
class recall	100.00%	8.91%	

Gambar 15. Confusion matrix

Berikut ialah yang akan terjadi akhir asal implementasi prosedur pemecahan *Naïve Bayes confusion matrix*, yang akan terjadi dari data yang pada bagi sebagai data latih serta data uji penjabaran memakai prosedur pemecahan naïve bayes pada rapidminer. menerima nilai akurasi sebanyak 67,13%. lalu di dapat yang akan terjadi precision sebanyak 66,04%, serta recall sebesar 100.00%.

4. KESIMPULAN

Berdasarkan pengolahan data *Twitter* sebanyak 3057 data, dengan topik “KPU 2024”. Hasil yang diperoleh hanya 715 data, sebab di waktu pencucian cara sebelumnya, dilakukan penghapusan istilah-istilah yang tidak digunakan. Selanjutnya juga dilakukan penghapusan kalimat yang ganda oleh sebab itu, total akhir data yang diperoleh hannya 715 data. Sebelum ketahap pelabelan secara otomatis, maka dibutuhkan hadiah label lebih-lebih dulu secara manual sebanyak 300 data oleh peneliti dan didapat sentimen positif 210 dan sentimen negative 90, proses ini dilakukan supaya *naïve*

bayes bisa melatih dirinya dari pelabelan yang telah dilaksanakan sebelumnya. Makin poly data train yang penelitian labeli secara manual, semakin seksama pula bagi *naïve bayes* di proses pelabelan secara otomatis nanti. Hasil akurasi yang diberikan oleh algoritma *naïve bayes* adalah sebesar 67,13% dengan precision sebesar 66,04%, dan recall sebesar 100.00%. Penelitian ini dilakukan untuk melihat sentimen masyarakat terhadap “KPU 2024” nantinya. Hasil evaluasi dalam *confusion matrix* didapat true positif sebesar 457 dan true negative sebesar 235.

REFERENCES

- [1] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [2] A. Rahma and S. Delliana, “Pengaruh Konten Youtube Ria Ricis Terhadap Agresi Verbal Netizen,” *Translitera J. Kaji.*, vol. 11, no. 1, pp. 106–119, 2022, doi: <https://doi.org/10.35457/translitera.v11i1.1668>.
- [3] A. R. Abdillah, F. N. Hasan, T. Informatika, F. N. Hasan, D. Mining, and A. Sentimen, “Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier Sentiment Analysis of Presidential Candidates Based on Tweets on,” vol. 13, no. 1, pp. 117–130, 2023, doi: <https://doi.org/10.32664/smatika.v13i01.750>.
- [4] W. Maulana and M. Mulyadi, “Ujaran Kebencian Terhadap Jokowi pada Masa Pandemi Covid-19: Studi Kasus Twitter,” *J. Lisnguistik Komputasional*, vol. 4, no. 1, pp. 27–33, 2021, doi: <https://doi.org/10.26418/jlk.v4i1.42>.
- [5] A. R. Alifvia, U. Saprudin, K. Metro, and R. Miner, “ANALISIS SENTIMEN REVIEW DATA TWITTER KOMISI PEMILIHAN UMUM (KPU) MENGGUNAKAN METODE NAÏVE BAYES,” vol. 11, no. 1, pp. 81–84, 2023, doi: <https://doi.org/10.35959/jik.v11i01.407>.
- [6] B. J. Yuri, A. Zetra, and R. E. Putera, “Analisis Kebijakan KPU dalam Rekrutmen KPPS untuk Pemilu 2024,” *SOSIOHUMANIORA J. Ilm. Ilmu Sos. Dan Hum.*, vol. 9, no. 1, pp. 145–154, 2023, doi: 10.30738/sosio.v9i1.13954.
- [7] R. Ambarwati, Aidinil Zetra, and Syahrizal, “Efektivitas Sosialisasi Pemilu 2019 KPU Kabupaten Indragiri Hulu,” *J. Niara*, vol. 15, no. 3, pp. 571–580, 2023, doi: 10.31849/niara.v15i3.9280.
- [8] M. Renaldi, U. Juhardi, A. Munandar, A. Karolina, and W. Seprianti, “Manajemen Pengarsipan Jaringan Dokumentasi Dan Informasi Hukum (Jdih) Kpu Provisi Bengkulu,” *J. Ilm. Mhs. Kuliah Kerja Nyata*, vol. 2, no. 2, pp. 224–230, 2022, doi: 10.36085/jimakukerta.v2i2.3390.
- [9] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.
- [10] P. Arsi, R. Wahyudi, and R. Waluyo, “Optimasi SVM Berbasis PSO pada Analisis Sentimen Wacana Pindah Ibu Kota Indonesia,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 231–237, 2021, doi: 10.29207/resti.v5i2.2698.
- [11] Alfandi Safira and F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [12] F. N. Hasan and M. Dwijayanti, “Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier,” *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [13] W. Khoffifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 171–180, 2022, doi: 10.35969/interkom.v16i4.192.
- [14] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier,” *J. Linguist. Komputasional*, vol. 5, no. 1, pp. 34–43, 2022, doi: 10.26418/jlk.v5i1.79.
- [15] E. Fitriani, “Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi*, vol. 9, no. 1, pp. 103–116, 2020, doi: 10.32520/stmsi.v9i1.596.
- [16] T. Astuti and Y. Astuti, “Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO),” *J. Media Inform. Budidarma*, vol. 6, no. 4, pp. 1806–1815, 2022, doi: 10.30865/mib.v6i4.4119.
- [17] N. Luh *et al.*, “ANALISIS SENTIMEN ULASAN VILLA DI UBUD MENGGUNAKAN METODE NAÏVE BAYES , DECISION TREE , DAN K-NN Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI | 206,” vol. 11, no. 3, pp. 205–216, 2022, doi: <https://doi.org/10.23887/janapati.v11i3.49450>.
- [18] N. Q. Rizkina and F. N. Hasan, “Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes,” vol. 4, no. 4, pp. 1136–1144, 2023, doi: 10.47065/josh.v4i4.3803.
- [19] I. T. Julianto, “Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut,” *J. Algoritma.*, vol. 19, no. 1, pp. 458–465, 2022, doi: 10.33364/algoritma/v.19-1.1112.
- [20] D. Hermawan, M. Akhsanal, Z. Wahyudi, A. Ariyanto, and D. Dwi, “05. Optimasi Analisis Sentimen Pada Twitter Olshop,” vol. 6, no. September, pp. 821–828, 2022, doi: <http://dx.doi.org/10.30645/j-sakti.v6i2.493>.
- [21] I. R. Afandi, I. F. Hanif, F. N. Hasan, E. Sinduningrum, Z. Halim, and N. Pratiwi, “Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression,” *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 77–84, 2022, doi: <https://doi.org/10.26418/jlk.v5i2.103>.