

ANALISIS SENTIMEN PENERAPAN KURIKULUM MERDEKA PADA TWITTER DENGAN METODE NAÏVE BAYES

Akbar Adi Susanto^{1*}, Painem², Mohammad Syafrullah³, Rizky Pradana⁴

^{1,2,3} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

⁴ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia

Email: ^{1*}akbaraadis@gmail.com, ²painem@budiluhur.ac.id, ³mohammad.syafrullah@budiluhur.ac.id,
⁴rizky.pradana@budiluhur.ac.id

Abstrak- Bulan Februari tahun 2022 Kemendikbudristek baru saja meluncurkan kurikulum terbaru yang dinamakan dengan nama Kurikulum Merdeka. Kurikulum Merdeka merupakan kurikulum pembelajaran intrakurikuler yang beragam, dimana konten akan lebih optimal agar peserta didik mempunyai waktu yang cukup untuk mendalami konsep dan menguatkan kompetensi. Salah satu tujuan dari Kurikulum Merdeka yaitu menciptakan pendidikan yang menyenangkan bagi peserta didik dan guru. Nantinya, guru memiliki kekuasaan untuk memilih berbagai macam perangkat ajar sehingga pembelajaran bisa disesuaikan dengan kebutuhan belajar dan minat peserta didik. Kurikulum ini untuk menguatkan pencapaian profil pelajar Pancasila yang dikembangkan berdasarkan tema tertentu oleh pemerintah. Yang mana proyek tersebut tidak diarahkan untuk mencapai target capaian pembelajaran tertentu, sehingga tidak terikat pada konten mata pelajaran. Dalam penelitian ini membahas mengenai kurikulum tersebut dengan tujuan untuk melakukan klasifikasi dari tweet yang didapatkan melalui API Twitter mengenai penerapan dari Kurikulum Merdeka dengan metode yang digunakan adalah Multinomial Naïve Bayes. Terdapat beberapa tahapan klasifikasi yang dilakukan yaitu pengumpulan data, pelabelan data, preprocessing, klasifikasi, dan pengujian. Data yang terdapat dalam penelitian sejumlah 538 data yang dikumpulkan dengan periode waktu 5 November sampai dengan 14 November 2022 dan hanya digunakan untuk klasifikasi sebanyak 256 dengan data tweet positif berjumlah 135 dan negatif berjumlah 121. Dan hasil pengujiannya menunjukkan nilai akurasi yang didapatkan sebesar 96%, kemudian presisi 100%, dan recall 95%.

Kata Kunci: Kurikulum Merdeka, *Multinomial Naïve Bayes*, Sentimen Analisis.

SENTIMENT ANALYSIS OF IMPLEMENTING INDEPENDENCE CURRICULUM ON TWITTER USING THE NAÏVE BAYES METHOD

Abstract- In February 2022 the Ministry of Education and Culture has just launched a new curriculum called the Independent Curriculum. The Merdeka Curriculum is a diverse extracurricular learning curriculum, where the content will be more optimal so that students have enough time to explore concepts and strengthen competence. One of the goals of the Merdeka Curriculum is to create fun education for students and teachers. Later, the teacher has the power to choose various kinds of teaching tools so that learning can be adapted to the learning needs and interests of students. This curriculum is to strengthen the achievement of Pancasila student profiles which were developed based on certain themes by the government. In which the project is not directed to achieve certain learning achievement targets, so it is not tied to subject content. This study discusses the curriculum to classify tweets obtained through the Twitter API regarding the implementation of the Merdeka Curriculum with the method used is Multinomial Naïve Bayes. There are several stages of classification carried out, namely data collection, data labeling, preprocessing, classification, and testing. The data contained in the study amounted to 538 data collected for the period 5 November to 14 November 2022 and was only used for 256 classifications with 135 positive tweets and 121 negative tweets. And the test results showed an accuracy value of 96%, then 100% precision, and 95% recall.

Keywords: Independent Curriculum, *Multinomial Naïve Bayes*, Sentiment Analysis.

1. PENDAHULUAN

Pada dasarnya manusia memiliki beberapa hak yang telah dimiliki sejak lahir, salah satunya yaitu hak kebebasan berpendapat. Kebebasan berpendapat dapat diartikan sebagai hak yang melekat dalam diri seseorang seperti memiliki dan menyampaikan pendapatnya. Saat ini dalam menyampaikan pendapat sudah dapat dilakukan dengan berbagai cara seperti menyampaikan pendapat secara langsung maupun tidak langsung. Untuk penyampaian pendapat secara tidak langsung telah difasilitasi oleh sebuah platform digital dengan logo burung berwarna biru atau biasa dikenal dengan nama Twitter.

Dikutip dari [1] bahwa terdapat setidaknya 18,45 juta pengguna Twitter di Indonesia pada tahun 2022, meningkat dari jumlah pengguna Twitter di tahun 2021 yakni hanya 14,50 juta. Dengan jumlah pengguna tersebut Indonesia menempati sebagai negara pengguna Twitter terbanyak ke 5 di dunia, adapun [2] memaparkan salah

satu alasan pengguna Twitter di Indonesia semakin tinggi yaitu karena banyak digunakan oleh para politikus dalam negeri dan mancanegara untuk mempromosikan kebijakan dan berinteraksi dengan warga ataupun pejabat lainnya. Kemendikbudristek baru saja meluncurkan kurikulum terbaru pendidikan Indonesia yaitu Kurikulum Merdeka yang baru diluncurkan sejak Februari 2022 sebagai salah satu program Merdeka Belajar.

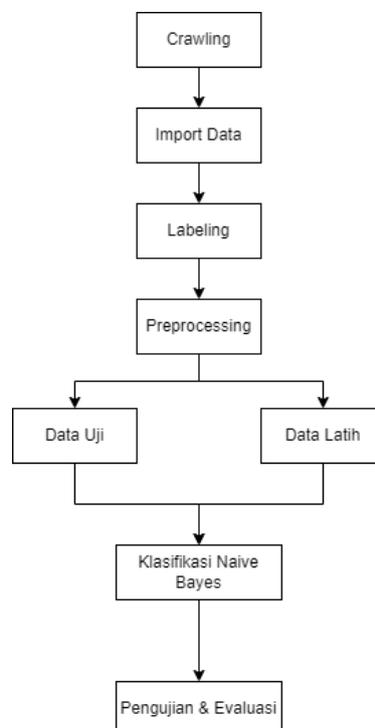
Kurikulum Merdeka adalah silabus kegiatan pembelajaran yang beragam, dengan memanfaatkan penggunaan konten sehingga pengajar memiliki opsi untuk memilih berbagai macam perangkat pembelajaran yang dapat disesuaikan dengan kegiatan belajar dan ketertarikan peserta didik. Dengan diterapkannya kurikulum baru tersebut terdapat berbagai macam pendapat terkait dengan kebijakan yang dilaksanakan, ada yang mendukung dan ada juga yang keberatan. Masing masing memiliki pendapatnya dari sudut pandang yang berbeda, melihat dari permasalahan tersebut peneliti akan melakukan analisis sentimen untuk mengevaluasi kebijakan pemerintahan yang diperbincangkan di Twitter.

Hal tersebut diharapkan dapat menjadi salah satu masukan penilaian opini publik terkait implementasi kurikulum merdeka agar dapat dikembangkan menjadi lebih baik lagi. Dengan melihat permasalahan tersebut, peneliti akan melakukan analisis sentiment untuk mengevaluasi kebijakan pemerintahan yang diperbincangkan di media sosial Twitter. Dalam [3] analisis sentiment dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seorang, apakah cenderung berpandangan atau beropini negative atau positif. Salah satu algoritma klasifikasi data dengan probabilitas objek di masa depan dan telah dilakukan sebuah penelitian oleh [4] didapatkan hasil akurasi yang cukup baik yakni sebesar 94,33% dari 1007 jumlah data tweet yang dipergunakan yaitu algoritma *Multinomial Naive Bayes*. Sehingga pada penelitian ini, bertujuan untuk mengembangkan sebuah program yang mampu melakukan analisis sentimen dengan mengimplementasikan algoritma tersebut dengan menggunakan dataset berupa data *tweet*, yang bertujuan untuk melakukan analisis sentimen atau pandangan masyarakat terhadap kebijakan Kurikulum Merdeka.

2. METODE PENELITIAN

2.1 Penerapan Metode

Dalam penelitian ini, terdapat alur atau tahapan yang digunakan dalam penelitian ini dengan menggambarkan tahapan awal hingga akhir. Dengan tujuan agar penelitian ini dapat berjalan sesuai dengan rencana. Penerapan metode terdapat pada Gambar 1.



Gambar 1. Penerapan Metode

Tahapan pertama yaitu pengumpulan data atau yang biasa disebut dengan tahapan *Crawling*. Referensi [5] menjelaskan bahwa Metode *Crawling* memanfaatkan API (*Application Programming Interface*) dalam pengambilan data sebagai datasetnya. Dalam penelitian ini, dilakukan crawling data tweet pada media sosial twitter

dengan menggunakan bahasa pemrograman Python menggunakan library Tweepy serta memanfaatkan Twitter API. Crawling data tweet dilakukan berdasarkan tweet yang memiliki kata kunci “kurikulum merdeka”, kemudian hasil dari data yang telah dikumpulkan akan disimpan di sebuah file berekstensi (.csv), yang kemudian dimasukkan ke dalam basis Data (Dataabase) MySQL. Kemudian untuk tahapan kedua, hasil data dikategorikan sebagai dirty dataset mentah. Kemudian pada tahap ketiga yaitu labeling, dataset yang diinput ke dalam dirt dataset akan diberikan 2 kategori label yaitu positif dan negatif dalam rangka kebutuhan training data dan menjadi acuan dalam penentuan akurasi dan klasifikasi.

Pada tahap selanjutnya yaitu preprocessing, dilakukan dengan tujuan untuk memudahkan komputer dalam mengenali datanya sehingga akan di dapatkan hasil yang semakin baik seperti yang dipaparkan oleh [6]. Terdapat beberapa proses untuk mendapatkan clean dataset, sehingga proses pengujian hasil klasifikasi sentimen lebih tepat dan akurat seperti (*case folding, cleansing, slangword, stopword, stemming*). Kemudian pada tahap *count vectorizer* bertujuan untuk mencari fitur kata dari data latih. Referensi [7] menunjukkan CountVectorizer, berfungsi untuk menghitung frekuensi kata dalam dokumen dan dapat mengubah fitur teks menjadi sebuah representasi vector. Setelah fitur kata telah didapatkan, dilakukan penghitungan setiap kemunculan fitur kata tersebut dari data latih pada masing-masing label.

Hasil perhitungan kemunculan fitur kata tersebut nantinya akan digunakan pada tahapan berikutnya yaitu tahapan klasifikasi. Kemudian tahap klasifikasi yang digunakan adalah Naïve Bayes, dalam [8] Bayesian classifiers mempunyai tingkat kecepatan dan akurasi yang tinggi ketika diaplikasikan dalam database yang besar kemudian didukung pula oleh [9] yang mengatakan bahwa kelebihan dari Naïve Bayes adalah sederhana dan berakurasi tinggi. Kemudian yang terakhir dilakukan pengujian melalui *confusion matrix* yang digambarkan seperti gambar berikut.

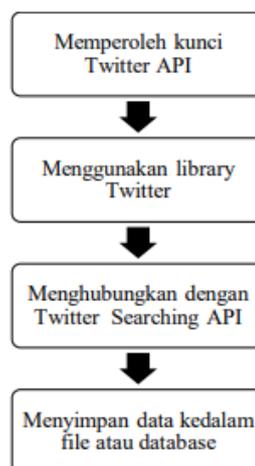
Tabel 1. Confusion Matrix

Correct Classification	Classified as	
	Predicted “+”	Predicted “-“
Actual “+”	True Positives	False Negatives
Actual “-“	False Positives	True Negatives

Confusion matrix ini akan digunakan untuk membandingkan data hasil prediksi (data dengan label dari hasil klasifikasi menggunakan Algoritma *Multinomial Naïve Bayes*) dengan data aktual (data dengan label dari tahap pelabelan secara manual). Setelah membandingkan kedua data tersebut, kemudian pengujian dilakukan untuk mengetahui nilai akurasi, presisi, dan *recall*.

2.2 Data Penelitian

Data yang digunakan dalam penelitian ini diperoleh dengan memanfaatkan proses *crawling* yaitu proses pengambilan data yang tersedia secara *online* dengan memindahkan informasi ataupun data yang didapatkan berdasarkan perintah tertentu ke file lokal di dalam komputer. Dan pada penelitian ini teknik *crawling* digunakan pada media sosial Twitter untuk mendapatkan sentimen dari user-user Twitter, seperti yang ditunjukkan pada gambar 2.



Gambar 2. Proses Mendapatkan Data

Berikut ini adalah penjelasan dari gambar 2 :

a. Memperoleh kunci Twitter API

Langkah pertama yang perlu dilakukan yakni membuat akun Twitter Developer untuk memperoleh *API Key*, *API Secret*, *Access Token*, dan *Access Token Secret*. Untuk lebih detailnya berikut ini adalah tahapannya.

- Mendaftarkan akun di Twitter, apabila belum memiliki
- Membuka situs <https://apps.twitter.com/> dan melakukan *log in* dengan akun yang sudah didaftarkan
- Menekan “*Create New App*” untuk membuat aplikasi baru
- Melengkapi form yang diberikan dan menyetujui term yang ada, kemudian menekan “*Create Your Twitter Application*”
- Pada halaman berikutnya, klik tab “*Keys and Access Tokens*”, pengguna dapat melihat *API Key*, *API Secret*, *Access Token* dan *Access Token Secret*, dan dapat memanfaatkannya pada aplikasinya.

b. Menggunakan library Twitter

Dari aplikasi yang dikembangkan atau dapat juga menggunakan text editor seperti visual studio code, ataupun google colab, melakukan *import library* Twitter untuk dapat menggunakan fitur-fitur yang tersambung secara langsung dengan Twitter.

c. Menghubungkan dengan Twitter Searching API

Tahapan berikutnya yaitu menghubungkan aplikasi dengan Twitter dengan cara menambahkan *API Key*, *API Secret*, *Access Token* dan *Access Token Secret* dalam file pemrograman.

d. Menyimpan data ke dalam file atau database

Setelah aplikasi dijalankan dengan menggunakan kata kunci atau keyword “kurikulum merdeka” dengan rentang waktu 5 – 14 November 2022 dan diperoleh data sejumlah 538 data yang disimpan dalam file, berupa CSV ke dalam *database*.

Dari 538 data yang telah didapatkan, berikut ini adalah beberapa contoh *sample* dari data yang digunakan pada penelitian ini seperti yang ditampilkan pada tabel berikut.

Tabel 2. Sampel Data

ID	Username	Tweet
1592101945313220000	mawar	KURIKULUM MERDEKA GAK ASIK BGT SETIAP HARI KELOMPOK DULU AWAL MPLS BANYAK KERJA KELOMPOK GT, MALESNYA KALO KELOMPOK TUH ADA YANG CUMAN NUMPANG NILAI SAMA GABISA ISTIRAHAT PAS WAKTU LIBUR. TOLONG DONG PLS APALAGI ADA P5 MANA SKOLAH GW NGADAINNYA TEMA DRAMA MALAH TAMBAH PUSING.'
1592057550597650000	j	cape banget sama project kurikulum merdeka
1591888308711610000	lexaa	ngang ngong ngang ngong tiba? dah mau project, abis project ada PAS?? ini bukan kurikulum merdeka anjr, ini kurikulum cobaan
1589124033072992257	Universitas Trisakti	Harapannya, sesuai dengan prinsip Kampus Merdeka yang mendekatkan Perguruan Tinggi dengan Praktisi, dapat dikembangkan kurikulum pendidikan tinggi yang sesuai dengan kebutuhan industri dan pengalaman praktisi energi.
1589430181504352256	ARIN????	@toorubreezyyg YAAMPUN ?? bener!! Apalagi kurikulum merdeka ada project T-T huhu semangat buat kita T-T? sipp kamu juga jaga kesehatan yaa <3

3. HASIL DAN PEMBAHASAN

3.1 Lingkungan Percobaan

Dalam penelitian ini, adapun spesifikasi dari perangkat yang digunakan untuk menjalankannya aplikasi ini sebagai berikut.

3.1.1 Spesifikasi Perangkat Keras dan Perangkat Lunak

Spesifikasi dari perangkat keras yang menjadi lingkungan percobaan dan menjalankan aplikasi dalam penelitian ini, yaitu:

- a. Komponen Processor : Intel i7-9750H

- b. Komponen RAM : 16 GB
- c. Komponen Storage : Harddisk 1 TB
- d. Komponen Display : 1920 x 1080

Berikut ini adalah spesifikasi perangkat lunak yang menjadi lingkungan percobaan dan menjalankan aplikasi dalam penelitian ini, yaitu:

- a. Sistem Operasi : Windows 10 Home
- b. Bahasa Pemrograman : Python 3.8.5
- c. IDE : Visual Studio Code v1.74.2
- d. DBMS : MySQL Database
- e. Browser : Microsoft Edge
- f. Lainnya : XAMPP v3.2.2

3.2 Implementasi Metode

Implementasi metode dalam penelitian ini mengacu kepada gambar 3.2 yang terdiri dari empat tahapan utama dimulai dari tahapan pengumpulan data (*crawling*), tahapan pelabelan (*labeling*), tahapan preprocessing, dan tahapan klasifikasi multinomial naïve bayes. Tahapan pengumpulan data atau dapat disebut juga tahapan *crawling* dimana tahapan ini merupakan tahapan pembukaan dalam penelitian ini. Data yang dikumpulkan didapatkan dari media sosial Twitter dengan memanfaatkan API Twitter yang telah disediakan dengan menggunakan bahasa pemrograman Python dan kata kuncinya yaitu “kurikulum merdeka”.

Dataset yang dikumpulkan mulai dari tanggal 5 sampai dengan 14 November 2022. Dan diperoleh dataset sebanyak 538 data dengan format (.csv) yang kemudian dimasukkan ke dalam database untuk dilakukan tahapan berikutnya. Kemudian, Tahapan pelabelan ini dilakukan untuk memberikan label atau kategori dari data tweet termasuk ke dalam jenis tweet yang positif atau negatif terkait sentimen terhadap kurikulum merdeka. Pada Tabel 3 menampilkan ulasan yang telah diberi label.

Tabel 3. Labeling

No	Tweet	Label
1	IH SINTING TUGAS GUE BANYAK AMAT DAH, kurikulum merdeka yg merdeka ya siapa ???	Negatif
2	@toorubreezyg YAAMPUN ?? bener!! Apalagi kurikulum merdeka ada project T-T huhu semangat buat kita T-T? sipp kamu juga jaga kesehatan yaa <3	Positif
3	kurikulum merdeka, merdeka drmn nya si ajg memo hp abis download pdf vidio bikin proposal naskah CAPEKKKKKKKK	Negatif
4	Enggak tau antara makin ribet atau emang belum terbiasa aja, ngeliat perangkat pembelajaran kurikulum merdeka belajar kok pusing ya ngeliatnya?! ?	Negatif
5	@Guru_PNS Semoga kurikulum merdeka benar benar memerdekakan guru Dr tugas tugas selain mengajar Dan mengembangkan diri	Positif
6	@tanyakanrl KURIKULUM MERDEKA TUH KEK ANJIR TAU. GAK CUMA SISWANYA YANG DI BIKIN STRES, GURUNYA JUGA. AWALNYA GUE PIKIR CUMA GUE AJA YANG STRES GEGARA KURIKULUM INI, TERNYATA KAKAK GUE NGELUH TEMEN TEMEN GUE PUN BANYAK YANG NGELUH BELUM LAGI DENGAR KELUHAN SISWA. GAK ADA MERDEKA MERDEKANYA	Negatif

Tahapan selanjutnya yaitu tahapan *preprocessing*, dalam [10] Tahapan preprocessing adalah tahapan untuk memproses data yang sebelumnya masih di *dataset* mentah untuk dilakukan proses pembersihan sehingga data menjadi lebih mudah untuk dikenali oleh sistem dalam metode text mining. Yang dalam penelitian terdiri atas 5 proses yaitu *case folding*, *cleansing*, *slangword*, *stopword*, *stemming*. Proses *case folding* yang berfungsi untuk merubah isi teks menjadi huruf kecil semua (*lower case*), selanjutnya merupakan proses *cleansing* yang berfungsi untuk menghapus karakter selain a-z, menghapus teks 1 karakter, menghapus URL, dan menghapus spasi berlebih. Tahapan berikutnya yaitu *slangword*, yang berfungsi untuk mengganti kata gaul menjadi kata baku, dan dalam tahapan ini dilakukan proses *stopword*, yang berfungsi untuk menghilangkan kata hubung atau kata sambung.

Berikutnya dalam tahapan *preprocessing* yang terakhir dilakukan proses *stemming*, yang berfungsi untuk merubah sebuah kata menjadi kata dasarnya. Setelah melalui berbagai tahapan sebelumnya kemudian akan dilakukan proses berikutnya dimana data yang telah bersih dari hasil *preprocessing* akan disimpan menjadi *clean*

data dan dilakukan pembagian menjadi data *training* dan data *testing* dengan rasio 9:1. Berikut ini adalah sample data training yang ditunjukkan pada table berikut.

Tabel 4. Data Training

No	Tweet	Label
1	salam semangat anak kurikulum merdeka anak kurtilas	positif
2	capek nder kurikulum merdeka tuh susah	negatif
3	iya ihhh kurikulum merdeka bikin stress	negatif
4	wkwkw semangat kurikulum merdeka	positif
5	perintah usaha baik mutu didik	positif
6	kurikulum merdeka tugas kelompok nya capek nder	negatif

Proses berikutnya adalah menemukan kata-kata unik terhadap kata yang terdapat di *training* data baik positif maupun negatifnya seperti yang dapat dilihat pada tabel berikut.

Tabel 5. Pencarian Kata Unik

Tweet Bersih	Kata	Kata Unik	Label
salam semangat anak kurikulum merdeka anak kurtilas	salam	salam	positif
	semangat	semangat	
	anak	anak	
	kurikulum	kurikulum	
	merdeka	merdeka	
	kurtilas	kurtilas	
capek nder kurikulum merdeka tuh susah	capek	capek	negatif
	nder	nder	
	kurikulum	-	
	merdeka	-	
	tuh	tuh	
	susah	susah	
iya ihhh kurikulum merdeka bikin stress	iya	iya	negatif
	ihhh	ihhh	
	kurikulum	-	
	merdeka	-	
	bikin	bikin	
	stress	stress	
wkwkw semangat kurikulum merdeka	wkwkw	wkwkw	positif
	semangat	-	
	kurikulum	-	
perintah usaha baik mutu didik	perintah	perintah	positif
	usaha	usaha	
	baik	baik	
	mutu	mutu	
	didik	didik	
	kurikulum	-	
kurikulum merdeka tugas kelompok nya capek nder	kurikulum	-	negatif
	merdeka	-	
	tugas	tugas	
	kelompok	kelompok	
	nya	nya	
	capek	-	
Total Kata Unik	nder	-	23

Kemudian dari total kata unik yang didapatkan dihitung kembali kemunculan kata uniknya pada masing-masing kategori. Setelah didapatkan total kemunculan kata pada masing-masing label kategori, nantinya total kata tersebut akan digunakan kembali pada tahap berikutnya. Dari *sample data training* yang digunakan di atas, terdapat 6 *sample data training*. Data dengan label positif berjumlah 3 data dan data dengan label negatif berjumlah 3 data juga. Untuk penghitungan probabilitas kelas menggunakan persamaan.

$$P(v_{\text{Positif}}) = \frac{|Data\ Positif|}{|Total\ Data\ Training|}$$

$$P(v_{\text{Positif}}) = \frac{|3|}{|6|}$$

$$P(v_{\text{Positif}}) = 0,5$$

$$P(v_{\text{Negatif}}) = \frac{|Data\ Negatif|}{|Total\ Data\ Training|}$$

$$P(v_{\text{Negatif}}) = \frac{|3|}{|6|}$$

$$P(v_{\text{Negatif}}) = 0,5$$

Tabel 6. Data Testing

No	Tweet	Label
1	capek banget project kurikulum merdeka	negatif

Tahapan berikutnya adalah klasifikasi dari *sample data testing*. Setelah melakukan perhitungan pada data *testing* terhadap masing-masing kategori label berikutnya adalah penentuan label dari data *testing*, dengan cara membandingkan hasil probabilitasnya dengan tiap label. Jika hasil probabilitas data *testing* terhadap kategori positif lebih besar dibandingkan dengan kategori negatif maka data tersebut akan diprediksi termasuk ke dalam kategori positif begitupun sebaliknya. Jika hasil probabilitas terhadap kategori negatif lebih besar dibandingkan dengan kategori positif, maka data tersebut akan diprediksi termasuk ke dalam kategori negatif. Dari proses perhitungan, diperoleh hasil probabilitas data *testing* terhadap kelas positif sebesar $4,96069561548e-8$ ($0,0000000496069561548$) sedangkan terhadap kelas negatif sebesar $1,83271961080e-7$ ($0,000000183271961080$). Dengan mengacu kepada kedua hasil tersebut dapat disimpulkan bahwa data *testing* akan diprediksi masuk ke dalam kategori negatif.

3.3 Analisa Pengajuan

Mengacu kepada hasil pengujian dengan perolehan akurasi, presisi, dan *recall* sebesar 96%, 100%, dan 95% terdapat beberapa faktor pendukung yang didapatkannya sebagai berikut.

a. Labeling

Labeling merupakan salah satu faktor pendukung dimana labeling ini mampu meningkatkan kualitas data pada suatu kategori, dan dijadikan acuan oleh sistem untuk melakukan proses data training.

b. Preprocessing

Kemudian preprocessing juga menjadi faktor pendukung dimana dalam tahapan ini dilakukan proses pembersihan kata-kata yang dianggap tidak dibutuhkan seperti url, emoticon, pergantian bahasa gaul atau kata tidak baku menjadi kata baku (slangword), dan lain sebagainya. Kemudian terdapat juga proses stemming untuk mengubah beberapa kata ber-imbunan menjadi kata dasar, hal ini sangat membantu mesin dalam mengolah data dan melakukan prediksi klasifikasi.

c. Penggunaan kata yang sama

Selain dari kedua faktor sebelumnya masih terdapat satu faktor pendukung lainnya yakni variasi atau penggunaan katanya tidak banyak. Dalam satu kategori tertentu setidaknya cukup banyak beberapa kata sama yang digunakan pada beberapa tweet sehingga proses perhitungan klasifikasi multinomial naïve bayesnya dapat lebih tinggi akurasi. Kata-kata tersebut seperti capek, stress, pusing, sakit, dan lain sebagainya.

4. KESIMPULAN

Berdasarkan hasil pengujian dari sistem analisis sentiment yang dibuat menggunakan dataset dan algoritma yang diusulkan, maka dapat disimpulkan bahwa berdasarkan 538 *tweet*, arah pandangan masyarakat mengenai penerapan kurikulum merdeka cenderung mengarah ke kategori positif dengan persentase 77,91% dengan rentan waktu 5 November – 14 November 2022. Tahapan preprocessing seperti *cleansing*, *slangword*, *stemming* sangat berpengaruh akan hasil yang didapatkan, semakin baik hasil preprocessing maka semakin akurat hasil prediksinya. Penggunaan metode *Multinomial Naïve Bayes* untuk analisis sentimen memberikan hasil persentase akurasi sebesar 96% dari hanya 256 data *tweet* yang digunakan dengan rasio pembagian 9:1 untuk data *training* dan data *testing*.

DAFTAR PUSTAKA

- [1] Hootsuite (We Are Social): Indonesian Digital Report 2matriz022. [Online]. Available: <https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2022>, 2022
- [2] A. Turmudi and K. S. Yasah, "Analisa Sentimen Tweet Indonesia Menggunakan Fitur Ekstraksi Dan Teknik Cross Validation Terhadap Model Naive Bayes," *SIGMA J. Teknol. Pelita Bangsa*, vol. 11, no. 3, pp. 137–142, 2020.
- [3] Rozi, I., Pramono, S. and Dahlan, E. "Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi," *Jurnal EECCIS*, vol. 6, no. 1, pp. 37-43, 2018.
- [4] P. Arsi, B. A. Kusuma, and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier," *J. Inform. Upgris*, vol. 7, no. 1, pp. 1–6, 2021, doi: 10.26877/jiu.v7i1.7636.
- [5] Kosasih, R. *et al.* "Perancangan Sistem Otomatis Transaksi Pembayaran Pada Marketplace UMKM Menggunakan Metode Crawling Horspool," vol. 6, pp. 2247-2256, 2022.
- [6] R. Naquitasia, D. H. Fudholi, and L. Iswari, "Analisis Sentimen Berbasis Aspek pada Wisata Halal dengan Metode Deep Learning," *J. Teknoinfo*, vol. 16, no. 2, p. 156, 2022, doi: 10.33365/jti.v16i2.1516.
- [7] Silitonga, Y. R. *et al.* "Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn," vol. 4, pp. 173-179, 2019.
- [8] Rahmah, A., Santoso, H. B. and Hasibuan, Z. A. "Characteristics analysis for technology enhanced learning maturity: A qualitative approach," *ICCE 2019 - 27th International Conference on Computers in Education, Proceedings*, vol. 1, no. 1, pp. 487-489, 2019.
- [9] H. F. Putro, R. T. Wulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [10] Nofiyani and Wulandari. "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *Jurnal Media Informatika Budidarma*, vol. 6, pp. 1621-1629, 2022. Nofiyani and Wulandari. "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *Jurnal Media Informatika Budidarma*, vol. 6, pp. 1621-1629, 2022.