

ANALISIS SENTIMEN KOMENTAR YOUTUBE PROGRAM KAMPUS MERDEKA BERBASIS WEB MENGUNAKAN ALGORITMA *MULTINOMIAL NAÏVE BAYES*

Rubi Ahmad Fauzan^{1*}, Mufti²

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

Email: ^{1*}1911501656@student.budiluhur.ac.id, ²mufti_hayat@yahoo.com

(* : corresponding author)

Abstrak- Program Kampus Merdeka yang diselenggarakan oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia disosialisasikan ke berbagai media salah satunya adalah sosial media *YouTube*. Dalam proses mengenalkan program tersebut, Kementerian Pendidikan dan Kebudayaan Republik Indonesia mengunggah *video* pada kanal resminya dengan judul “Merdeka Belajar Episode 2: Kampus Merdeka”. Disisi lain pada kanal *YouTube* Dr.IndrawanNugroho dan Opini Tempo juga membahas persoalan program Kampus Merdeka. *Video-video* tersebut mendapatkan berbagai sentiment positif dan negatif dengan memberikan komentar pada setiap *video* oleh masyarakat terutama pada mahasiswa maupun calon mahasiswa yang menjadi target utama dalam program tersebut. Data Komentar dari tiga *video* tersebut dapat dijadikan sebagai penelitian analisis sentimen terhadap komentar *YouTube* dengan menghasilkan dua kelas sentimen yaitu positif dan negatif. Analisis sentimen yang dilakukan tersebut menggunakan penerapan metode *Multinomial Naïve Bayes*. Algoritma Multinomial Naïve Bayes adalah metode pada text mining yang menggunakan asumsi bahwa setiap kata memiliki distribusi multinomial. Algoritma tersebut mempelajari frekuensi kata dalam setiap kelas kemudian memprediksi kelas untuk data baru berdasarkan dari kemungkinan kemunculan kata-kata tersebut. Hasil dari pengujian menggunakan metode Multinomial Naïve Bayes mendapatkan akurasi senilai 86.11% pada 359 dataset yang terbagi menjadi dua kelas yaitu 182 komentar positif dan 177 komentar negatif dengan dukungan tahapan *preprocessing* yang meliputi *case folding*, *cleaning*, *tokenizing*, *penggantian slang word*, penghapusan *stop word* dan *stemming*. Dari pengujian yang telah dilakukan menghasilkan kesimpulan bahwa metode *Multinomial Naïve Bayes* dapat digunakan untuk melakukan analisis sentimen terhadap komentar YouTube dengan hasil yang cukup baik. Dengan penelitian ini juga memperkuat metode *multinomial naïve bayes* sebagai alat yang efektif dalam menganalisis data teks dengan tujuan mengukur sentimen. Selain itu, adanya sistem analisis sentimen terhadap komentar YouTube tentang program Kampus Merdeka dapat bermanfaat bagi mahasiswa maupun calon mahasiswa yang menjadi peran utama dalam pelaksanaan program tersebut serta bahan evaluasi bagi Kementerian Pendidikan dan Kebudayaan Republik Indonesia dalam meningkatkan kualitas program.

Kata Kunci: *Sentiment Analysis, Multinomial Naïve Bayes, Kampus Merdeka, Komentar YouTube*

SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ABOUT PROGRAM USING NAÏVE BAYES MULTINOMIAL ALGORITHM

Abstract- The Kampus Merdeka Program organized by the Ministry of Education and Culture of the Republic of Indonesia is disseminated to various media, one of which is *YouTube* social media. In the process of introducing the program, the Ministry of Education and Culture of the Republic of Indonesia uploaded a video on its official channel with the title “Merdeka Belajar Episode 2: Kampus Merdeka”. On the other hand on the *YouTube* channel Dr.IndrawanNugroho and Opini Tempo also discussed the issue of the Kampus Merdeka Program. These videos received various positive and negative sentiments by providing comments on each video by the public, especially students and prospective students who were the main targets of the program. Comment data from the three videos can be used as sentiment analysis research on *YouTube* comments by producing two classes of positive and negative sentiments. Analysis This sentiment is carried out using the application of the *Multinomial Naïve Bayes* method. The *Multinomial Naïve Bayes* Algorithm is a method in text mining that uses the assumption that each word has a multinomial distribution. The algorithm studies the frequency of words in each class and then predicts the class for new data based on the probability of the occurrence of these words. The results of testing using the *Multinomial Naïve Bayes* method obtained an accuracy of 86.11% on 359 datasets which were divided into two classes, namely 182 positive comments and 177 negative comments with the support of preprocessing stages which included *case folding*, *cleaning*, *tokenizing*, replacing slang words, removing stop words and *stemming*. From the tests that have been carried out, it is concluded that the *Multinomial Naïve Bayes* method can be used to carry out sentiment analysis of *YouTube* comments with quite good results. This research also strengthens the *naïve Bayes multinomial* method as an effective tool in analyzing text data with the aim of measuring sentiment. In addition, the existence of a sentiment analysis system for *YouTube* comments about the Kampus Merdeka program, it can be useful for students and prospective students who play a major role in implementing the program as well as evaluation material for the Ministry of Education and Culture of the Republic of Indonesia in improving the quality of the program.

Keywords: Sentiment Analysis, Multinomial Naïve Bayes, Kampus Merdeka, YouTube Comment

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi memberikan kemudahan dalam berbagai bidang, salah satunya adalah bidang pendidikan, termasuk pendidikan tinggi. Mahasiswa sebagai salah satu pemegang perubahan perlu kemampuan untuk menyesuaikan diri dengan perubahan teknologi dan industri untuk terus mengembangkan diri. Dalam *video YouTube* yang diunggah oleh kanal KEMENDIKBUD RI, Nadiem Makarim selaku Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Indonesia yang menjabat sejak 28 April 2021, menjelaskan bahwa perguruan tinggi ditantang untuk menjawab kebutuhan industri yang selalu berubah, masih banyak kurikulum dari program studi pada perguruan tinggi memiliki sifat yang teoritis dan tidak 100% *link and match* dengan kebutuhan di dalam dunia nyata, meskipun materi yang ada pada perguruan tinggi bisa dibilang bagus, namun hal ini masih belum bisa bersaing di kancah internasional [1]. Untuk itu, Kementerian Pendidikan dan Kebudayaan Republik Indonesia menyelenggarakan program yang dinamakan Kampus Merdeka sebagai pendukung mahasiswa dalam mengembangkan diri.

Adanya program kampus merdeka yang kemudian disosialisasikan oleh Kementrian Pendidikan dan Kebudayaan Republik Indonesia di media sosial *YouTube* menuai berbagai sentimen dari masyarakat, terutama bagi mahasiswa dan calon mahasiswa, yang merupakan target utama dari program tersebut, sampai saat ini pada tanggal 30 Mei 2023 tercatat sebanyak 545 komentar. Selain pada *video* yang diunggah oleh kanal KEMENDIBUDRI, juga terdapat dua *video* serupa yang membahas tentang program kampus merdeka diantaranya adalah *video* yang pertama diunggah oleh Dr.IndrawanNugroho yang berjudul “Kampus Merdeka: Polemik dan Manfaatnya” yang dihadiri oleh Profesor Nizam dari Dirjen Dikti RI yang tercatat pada tanggal 30 Mei 2023 memiliki 343 komentar, kemudian *video* yang kedua diunggah oleh kanal media Opini Tempo dengan judul “Salah Kaprah Kampus Merdeka ala Nadiem Makarim” yang tercatat pada tanggal 30 Mei 2023 memiliki 154 komentar. Hal tersebut menarik penulis untuk melakukan penelitian untuk membuat sistem yang mendeteksi analisis sentimen masyarakat pengguna *YouTube* dalam menanggapi program Kampus Merdeka menggunakan algoritma *Multinomial Naïve Bayes* dengan judul “Analisis Sentimen Komentar *YouTube* Tentang Program Kampus Merdeka Berbasis Web Menggunakan Algoritma *Multinomial Naïve Bayes*”. Analisis tersebut menghasilkan 2 polaritas yaitu sentimen positif dan sentimen negatif.

Dalam beberapa penelitian sebelumnya telah membuat analisis sentimen menggunakan beberapa algoritma yang memiliki fungsi sama sebagai klasifikasi sentimen dan mencapai akurasi sebesar 79% dengan menggunakan metode *Random Forest* dan *F1-Score* sebesar 86,25% [2]. Kemudian analisis sentimen pengguna transportasi *online* terhadap layanan Grab Indonesia menggunakan metode *Multinomial Naïve Bayes Classifier* dengan hasil akurasi sebesar 86,57% [3]. Kemudian pada penelitian lain yang menggunakan metode metode *Naïve Bayes – Support Vector Machine (NBSVM) Classifier* dengan hasil kombinasi dari metode *Naïve Bayes* dan *Support Vector Machine* mencapai tingkat akurasi yang lebih baik dan performa yang lebih kuat dengan penggunaan skala data 7:3, yaitu 70% data *training* dan 30% data *testing*. Dengan menghasilkan nilai uji performansi terbesar yaitu *precision* sebesar 91%, *recall* sebesar 83%, dan *f1score* sebesar 87%[4]. Dan penelitian terhadap kebijakan kampus merdeka menggunakan metode *Naïve Bayes* dan pembobotan TF-IDF berdasarkan komentar pada *YouTube* dengan hasil akurasi terbaik sebesar 97% yang diperoleh dengan menggunakan 900 data latih, 100 data uji, menerapkan pembobotan TF-IDF, dan *10-fold cross validation*. Rata-rata akurasi yang didapat dari 10 literasi pada *k-fold cross validation* yaitu sebesar 91.8% dengan nilai *precision*, *recall*, *f-measure* sebesar 90.35%, 93.6%, 91.95% [5].

Beragam reaksi opini dalam kolom komentar pengguna ketika melihat konten *video* yang diunggah di *YouTube* sangat mempengaruhi reputasi konten *video* dan *channel* tersebut. Pendekatan *text mining* menjadi alternatif untuk mengartikan makna dari setiap komentar. Proses pengklasifikasian menjadi negatif dan positif menjadi sangat penting bagi pengguna *YouTube* untuk menilai seberapa bermakna konten yang telah diterbitkan berdasarkan opini pengguna. Hal tersebut menarik penulis untuk melakukan penelitian dalam membuat sistem yang mendeteksi analisis sentimen masyarakat pengguna *YouTube* dalam menanggapi program Kampus Merdeka menggunakan algoritma *Multinomial Naïve Bayes* dengan judul “Analisis Sentimen Komentar *YouTube* Tentang Program Kampus Merdeka Berbasis Web Menggunakan Algoritma *Multinomial Naïve Bayes*”. Analisis tersebut menghasilkan 2 polaritas yaitu sentimen positif dan sentimen negatif.

2. METODE PENELITIAN

2.1 Sentiment Analysis

Sentiment Analysis atau analisis sentimen didefinisikan sebagai ilmu yang menganalisis suatu pendapat, emosi, perasaan dan sikap seseorang yang terkandung dalam teks tertulis. Analisis sentimen merupakan proses untuk menganalisis dan mengelompokkan suatu kata yang dibagi menjadi beberapa kelas. Tujuan dari analisis sentimen itu sendiri adalah untuk mengekstrak atribut dan komponen dari beberapa komentar yang ada di media sosial, sehingga dapat menentukan beberapa kelas yang dibagi menjadi positif, negatif dan netral. [6]

Analisis sentimen dipergunakan untuk mendapatkan informasi penting dari data yang tidak terstruktur. Manfaat utama dari analisis sentimen adalah sebagai suatu cara efektif untuk mengidentifikasi dan mengklasifikasikan sikap dan sentimen pengguna (positif, negatif, dan netral) dalam teks untuk memastikan sikap *user* terhadap produk, subjek, atau layanan. [7]

2.2 Text Preprocessing

Pada tahap ini, data teks akan dibersihkan sehingga tersisa bentuk dasar dari masing-masing kata saja untuk keperluan analisis lebih lanjut. Terdapat beberapa tahapan yang dilakukan dalam fase *preprocessing*, sehingga data yang digunakan dapat siap untuk diolah pada fase berikutnya. Adapun beberapa tahapan dari *text processing* diantaranya adalah *case folding*, *cleaning*, *tokenizing*, *slang word replacement*, *stop word removal*, dan *stemming*. [8]

2.3 Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan salah satu metode spesifik dari metode *Naïve Bayes* sebagai metode *text mining* dalam proses klasifikasi teks berdasarkan probabilitas kelas dalam dokumen. *Multinomial Naïve Bayes* memperhitungkan jumlah kemunculan kata dalam dokumen, sehingga mengasumsikan independensi kehadiran kata pada manuskrip dengan tidak memperkirakan urutan kata ataupun konteks informasi. [9] Proses pembelajaran dimulai dari memasukkan data latih, kemudian dilakukan penghitungan prior/pelebaran kehadiran suatu kategori pada data latih dengan menggunakan Persamaan 1. [10]

$$P(c) = \frac{N_c}{N_{doc}} \quad (1)$$

Keterangan:

- c : Kelas
- doc : Dokumen
- N_c : Jumlah kategori c yang terdapat pada dokumen latih
- N_{doc} : Jumlah keseluruhan dari banyaknya dokumen latih yang digunakan

Kemudian langkah berikutnya adalah menghitung peluang pada sebuah kata i termasuk ke dalam kelas tertentu yang menggunakan Persamaan 2. [10]

$$P(w_i, c) = \frac{count(w_i, c) + 1}{\sum_w count(w, c) + |V|} \quad (2)$$

Keterangan:

- w_i : Posisi indeks ke- i di dalam semua dokumen yang dikategorikan sebagai c
- $count(w_i, c)$: Jumlah kemunculan kata-kata tertentu dalam suatu kategori.
- $\sum_w count(w, c)$: Total keseluruhan kata pada suatu kelas
- $|V|$: Total keseluruhan kata yang unik pada kelas

Penambahan nilai satu dimaksudkan agar tidak menghasilkan probabilitas bernilai 0, modifikasi ini dinamakan dengan *smoothing*. Setelah tahap pembelajaran selesai, tahap berikutnya adalah mengklasifikasikan data baru menggunakan hasil pembelajaran. Penghitungan klasifikasi data baru dengan digunakannya Persamaan 3. [10]

$$P(c, d) = P(c) \prod_{1 \leq i \leq count(v, d)} P(w_i, c) \quad (3)$$

Keterangan:

- $count(v, d)$: Total jumlah kata yang unik pada dokumen

2.4 Confussion Matrix

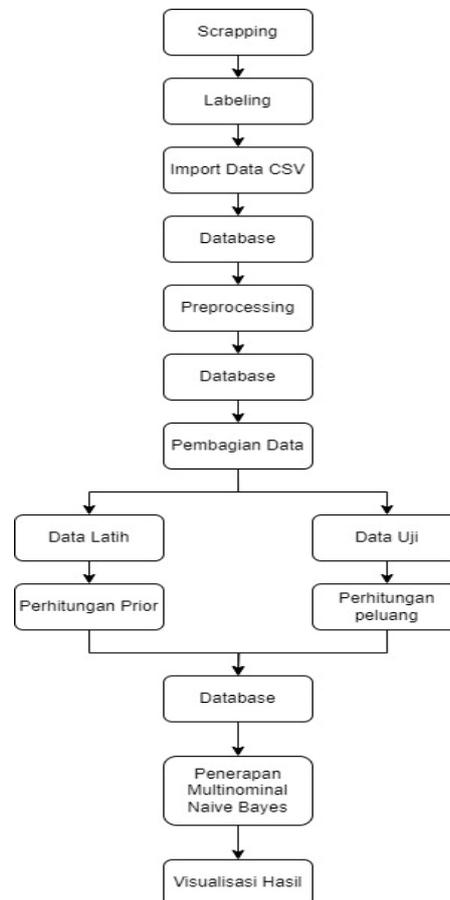
Confussion matrix merupakan suatu pendekatan yang bisa diterapkan untuk menghitung ketepatan dalam konsep penambangan data atau untuk mengukur sejauh mana kinerja proses klasifikasi. Menghitung akurasi menggunakan *confussion matrix* menghasilkan data dari model klasifikasi berupa *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN). Penghitungan akurasi dapat menggunakan Persamaan 4. [11]

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \times 100\% \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Metode

Gambar 1 adalah penjelasan alur tahapan metode pada penelitian ini.



Gambar 1. Tahapan Metode

Pada Gambar 1, langkah pertama yang dilakukan adalah scrapping menggunakan *YouTube API* yang kemudian diexport menjadi *file csv*. Pada *file csv* tersebut kemudian dilakukan pelabelan manual oleh pakar yang selanjutnya mengimpor data yang sudah diberi label sentimen positif dan negatif dalam bentuk *file* berformat *csv* yang memiliki *header text* dan *sentiment*. Kemudian data yang sudah terupload akan disimpan di *database*. Selanjutnya data tersebut melalui tahap *preprocessing* agar menghasilkan dataset yang bersih. Data kemudian dikonfirmasi dan akan disimpan ke *database* yang kemudian pada pembagian data menjadi data latih dan data uji yang memiliki perbandingan rasio 9:1. Data latih dilakukan perhitungan prior terhadap *training* data dan pada data

uji dilakukan perhitungan probabilitas pada sebuah kata dari setiap dokumen pada testing data yang selanjutnya melakukan klasifikasi agar menghasilkan sentiment *positive* atau *negative* terhadap data komentar yang kemudian di tampilkan hasilnya.

3.2 Scrapping

Tahapan pertama dilakukan *scrapping* untuk mendapatkan data menggunakan *YouTube API* dan *javascript* sederhana yang kemudian data tersebut di ekspor kedalam bentuk *file csv*.

3.3 Labeling

Tahapan selanjutnya dilakukan *labeling* manual terhadap data yang sudah diperoleh sebelumnya. Tahap *labeling* dilakukan oleh pakar dengan mengklasifikasikan setiap komentar terhadap dua kelas sentimen yaitu kelas positif dan kelas negatif. Pemberian label dilakukan menggunakan *Microsoft excel* yang kemudian disimpan menjadi *file csv*.

3.4 Import Data

Pada tahapan ini, mengimpor data komentar kedalam *database* yang sudah memiliki label *positive* dan *negative* dengan mengupload file berupa *csv* yang memiliki header text dan sentiment. Data komentar yang bersumber dari 3 *video* dari kanal *YouTube* yang berbeda, diantaranya kanal Dr.IndrawanNugroho, KEMENDIKBUDRI, dan Tempodotco.

3.5 Preprocessing

Preprocessing dilakukan pada setiap dokumen atau komentar untuk mendapatkan dataset bersih sehingga pengujian klasifikasi sentimen menghasilkan hasil yang akurat. Tahapan preprocessing meliputi:

a. *Case Folding*

Case folding adalah teknik mengubah sebuah kata menjadi huruf kecil seperti "Sentimen" menjadi "sentimen", "Analisis" menjadi "analisis" dan seterusnya.

b. *Cleaning*

Cleaning adalah proses membersihkan text menjadi bersih dengan menghapus karakter yang tidak relevan seperti tanda baca, simbol, atau URL.

c. *Tokenizing*

Tokenizing adalah teknik pemenggalan kalimat menjadi potongan kata untuk memudahkan proses selanjutnya. Proses ini melibatkan *library naturaljs*.

d. Mengganti *Slang Word*

Pada langkah ini, setiap kata dibandingkan dengan daftar kata *slang (slang word list)*. Untuk membandingkannya dengan kata yang telah ditokenisasi, penulis memasukan kata dalam bentuk JSON.

e. Menghapus *Stop Word*

Tahapan selanjutnya adalah menghapus *stop word* dari kata yang sudah diganti pada tahapan sebelumnya. Prosesnya sama seperti sebelumnya, data *stop word* dimasukkan kedalam *JSON*.

f. *Stemming*

Mengubah kata menjadi kata dasar dengan menghilangkan imbuhan seperti awalan, sisipan, akhiran, dan kombinasi antara awalan dan akhiran dikenal sebagai *stemming*. Misalnya kata "diatas" menjadi "atas" dan "membaca" menjadi "baca". Proses ini melibatkan *TS-Sastrawi Library*.

3.6 Multinomial Naïve Bayes

Tahapan berikut adalah menghitung menggunakan data sampel. Data sampel tersebut terbagi menjadi dua bagian antara lain data *training* dan data *testing*, hasil dari *preprocessing* terhadap sampel data latih dapat dilihat pada Tabel 1. Kemudian pada Tabel 2 ditunjukkan hasil dari *preprocessing* pada sampel data uji

Tabel 1 Hasil *Preprocessing* Pada Sampel Data Latih

No	Komentar	Label
1	mentor magang kampus merdeka syukur program usaha utama startup bantu selesai projek adik magang peras skill kreativitas program	Positif
2	keren program nya gabung kampus ajar	Positif
3	gebrak gabung kuliah dunia industri kuliah kerja kerja kuliah moga sukses program	Positif
4	nasib lulus bakal kena gusur bakal tinggal jauh gila sih lumayan susah saing dunia bisnis kerja upgrade skill kompetensi tambah sulit	Negatif

5	ganti menteri ganti program bakal program jangka panjang tri	Negatif
---	--	---------

Tabel 2 Hasil *Preprocessing* Pada Sampel Data Uji

No	Komentar	Label
1	keren banget program syukur banget mahasiswa	Positif

Pada tahapan selanjutnya, dilakukan perhitungan peluang kelas terhadap data latih serta data uji menggunakan persamaan pada Persamaan 1

$$P(\text{positif}) = \frac{3}{5}$$

$$P(\text{negatif}) = \frac{2}{5}$$

Selanjutnya dilakukan perhitungan probabilitas pada kata kata yang diambil dari data uji dengan menggunakan Persamaan 2.

Data Komentar : keren banget program syukur banget mahasiswa

Perhitungan kelas positif :

$$P(\text{keren, positif}) = \frac{1 + 1}{36 + 49} = \frac{2}{85}$$

$$P(\text{banget, positif}) = \frac{0 + 1}{36 + 49} = \frac{1}{85}$$

$$P(\text{program, positif}) = \frac{4 + 1}{36 + 49} = \frac{5}{85}$$

$$P(\text{syukur, positif}) = \frac{1 + 1}{36 + 49} = \frac{2}{85}$$

$$P(\text{banget, positif}) = \frac{0 + 1}{36 + 49} = \frac{1}{85}$$

$$P(\text{mahasiswa, positif}) = \frac{0 + 1}{36 + 49} = \frac{1}{85}$$

Perhitungan kelas negatif :

$$P(\text{keren, negatif}) = \frac{0 + 1}{30 + 49} = \frac{1}{79}$$

$$P(\text{banget, negatif}) = \frac{0 + 1}{30 + 49} = \frac{1}{79}$$

$$P(\text{program, negatif}) = \frac{2 + 1}{30 + 49} = \frac{3}{79}$$

$$P(\text{syukur, negatif}) = \frac{0 + 1}{30 + 49} = \frac{1}{79}$$

$$P(\text{banget, negatif}) = \frac{0 + 1}{30 + 49} = \frac{1}{79}$$

$$P(\text{mahasiswa, negatif}) = \frac{0 + 1}{30 + 49} = \frac{1}{79}$$

Langkah selanjutnya adalah melakukan hitungan probabilitas dokumen *testing* terhadap kelas sentimen menggunakan Persamaan 3

$$P(\text{positive, d1}) = \frac{3}{5} \times \left(\frac{2}{85}\right)^1 \times \left(\frac{1}{85}\right)^1 \times \left(\frac{5}{85}\right)^1 \times \left(\frac{2}{85}\right)^1 \times \left(\frac{1}{85}\right)^1 \times \left(\frac{1}{85}\right)^1$$

$$= 0.6 \times 0.023529 \times 0.011764 \times 0.058823$$

$$\times 0.023529 \times 0.011764 \times 0.011764$$

$$= 0.000000000318104$$

$$P(\text{negative, d1}) = \frac{2}{5} \times \left(\frac{1}{79}\right)^1 \times \left(\frac{1}{79}\right)^1 \times \left(\frac{3}{79}\right)^1 \times \left(\frac{1}{79}\right)^1 \times \left(\frac{1}{79}\right)^1 \times \left(\frac{1}{79}\right)^1$$

$$\begin{aligned}
 &= 0.4 \times 0.012658 \times 0.012658 \times 0.037974 \\
 &\quad \times 0.012658 \times 0.012658 \times 0.012658 \\
 &= 0.0000000000049359
 \end{aligned}$$

Hasil dari perhitungan probabilitas dokumen terhadap kelas tersebut menunjukkan bahwa kemungkinan dokumen terhadap kelas positif dengan nilai 0.000000000318104 yang lebih besar dibandingkan terhadap kelas negatif dengan nilai 0.000000000049359. Hal tersebut ditarik sebuah kesimpulan yang dimana sampel data training tersebut dikelompokkan atau diklasifikasikan kedalam *sentiment positive*.

3.7 Pengujian

Pengujian metode menggunakan dengan dataset dengan jumlah 359 komentar yang dapat dilihat pada Tabel 3 sebagai representasi data yang digunakan dalam pengujian metode.

Tabel 3 Data Pengujian

No	Komentar	Label
1	Sebagai mentor magang kampus merdeka saya sangat bersyukur sekali ada program ini, perusahaan (terutama startup) sangat terbantu dalam penyelesaian proyek dan adik2 magang jg sangat "diperas" skill dan kreativitasnya di program ini.	Positif
2	Keren program nya Saya bergabung di kampus mengajar 4	Positif
3	Bagaimana ini nasib yg cuma lulusan sma????, mungkin bakal kena gusur, bakal tertinggal jauh, gila sih mana udah lumayan susah buat bersaing di dunia bisnis sama kerja, ini lagi pada upgrade skill sama kompetensi, tambah sulit lagi.	Negatif
...
359	Tidak adanya jaringan lulusan di beberapa kampus yang akreditasi nya A. Mengapa seolah olah sekarang akreditasi pun tidak ada bedanya?	Negatif

Selanjutnya, hasil *confusion matrix* yang diperoleh dari pengujian dengan menggunakan pengaplikasian metode *multinomial naïve bayes* yang sudah dibuat dapat dilihat pada Tabel 4 yang terdiri dari *true positive* yang memiliki *actual label positive* dan *predicted label positive* dengan nilai 19, *true negative* yang memiliki *actual label negative* dan *predicted label negative*, *false positive* yang memiliki *actual label negative* dan *predicted label positive*, serta *false negative* yang memiliki *actual label positive* dan *predicted label negative*.

Tabel 4 Confusion Matrix

<i>True Positive</i>	<i>True Negative</i>	<i>False Positive</i>	<i>False Negative</i>
19	13	3	2

Kemudian menghitung akurasi berdasarkan Tabel 4 menggunakan Persamaan 4

$$\begin{aligned}
 Accuracy &= \frac{19 + 12}{19 + 12 + 3 + 2} \times 100\% \\
 &= \frac{31}{36} \times 100\% \\
 &= 86.11\%
 \end{aligned}$$

3.8 Analisa Hasil Pengujian

Berdasarkan hasil pengujian yang telah dilakukan, diperoleh akurasi yang tinggi yaitu 86,11%. Hasil pengujian dengan akurasi tinggi tersebut diperoleh dengan adanya beberapa faktor pendukung, antara lain:

a. *Labeling*

Tahapan ini digunakan sebagai salah satu faktor yang pendukung, dikarena tahapan *labeling* dilakukan penyeleksian data yang baik supaya membuat kualitas kamus data untuk kelas-kelas tertentu.

b. *Preprocessing*

Tahapan *text preprocessing* dijadikan sebagai faktor pendukung dikarenakan tahapan *preprocessing* dapat meningkatkan kualitas data dengan mengurangi redundansi kata dan bahkan kata yang tidak berguna dalam proses klasifikasi nantinya. Seperti contoh pada proses *stemming* dengan mengubah “mendengar” dan “dengarkan” menjadi kata dasar “dengar” yang dapat sehingga kata *dengar* memiliki 2 frekuensi.

4. KESIMPULAN

Hasil evaluasi dari sistem analisis sentimen komentar *YouTube* tentang program Kampus Merdeka, maka disimpulkan sebagai berikut:

- Sistem yang sudah dibuat dapat menganalisis sentimen terhadap komentar *YouTube* tentang program kampus merdeka.
- Penerapan metode *Multinomial Naïve Bayes* pada analisis sentimen menggunakan 359 dataset menghasilkan akurasi yang tinggi senilai 86,11 % yang terbagi menjadi dua bagian, yaitu data latih dan data uji dengan perbandingan antara data *training* dan data *testing* sebesar 9:1 serta perbandingan sentiment positif dan negatif adalah 1:1.
- Pemilihan data pada tahapan *labeling* berpengaruh penting dalam penentuan kualitas kamus kata mengenai kelas-kelas sentimen.
- Penggunaan *preprocessing* seperti mengganti *slang word* memiliki pengaruh dalam peningkatan frekuensi sehingga perhitungan peluang dapat ditingkatkan.

Sebagai bentuk pengembangan yang lebih lanjut agar sistem dapat berjalan lebih sempurna dan fungsi yang lebih baik, maka berikut ini adalah beberapa saran yang dapat penulis berikan:

- Pengkombinasian dalam pengembangan sistem dengan metode lain agar hasil yang didapatkan lebih baik.
- Penambahan fitur *scraping* otomatis.
- Penambahan kamus kata *slang word* yang lebih banyak agar pendeteksian *slang word* lebih baik.
- Penambahan kelas sentimen netral.

DAFTAR PUSTAKA

- [1] KEMENDIKBUDRI, “Merdeka Belajar Episode 2: ‘Kampus Merdeka,’” 2021. https://www.youtube.com/watch?v=xoQSIJSUUHl&ab_channel=KEMENDIKBUDRI
- [2] I. Afdhal, 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. 49–54, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- [3] S. Mandasari, B. H. Hayadi, and R. Gunawan, “Analisis Sentimen Pengguna Transportasi Online Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, p. 118, 2022, doi: 10.53513/jsk.v5i2.5635.
- [4] A. N. Muhamad, “Analisis Sentimen Positif Dan Negatif Komentar Video Youtube Menggunakan Metode Naïve Bayes - Support Vector Machine (Nbsvm) Classifier,” *Skripsi*, p. 17, 2019.
- [5] D. F. Zhafira, B. Rahayudi, and I. 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, doi: 10.25126/justi.v2i1.24.
- [6] P. A. Permatasari, L. Linawati, and L. Jasa, “Survei Tentang Analisis Sentimen Pada Media Sosial,” *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, p. 177, 2021, doi: 10.24843/mite.2021.v20i02.p01.
- [7] R. Y. Lesmana and R. Andarsyah, “Model Klasifikasi Multinomial Naïve Bayes Untuk Analisis Sentiment Terkait Non-Fungible Token,” *J. Tek. Inform.*, vol. 14, no. 3, pp. 135–139, 2022.
- [8] F. A. Nugraha, N. H. Harani, and R. Habib, *Analisa Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Cetakan 1. Bandung: Kreatif Industri Nusantara, pp. 2-6, 2020, ISBN: 978-623-7898-64-1.
- [9] A. A. Permana et al., *Machine Learning*. Padang: Global Eksekutif Teknologi, pp. 77, 2023, ISBN: 6231980723.
- [10] F. Hadaina and U. Budiyo, “Implementasi Metode Multinomial Naïve Bayes Untuk Sentiment Analysis Terhadap Data Ulasan Produk Colearn Pada Google Play Store,” *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 660–666, 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php>
- [11] Y. S. Hartini, A. B. P. Lefanska, A. A. Ursia, D. A. B. Prasetyo, and B. Sugiharto, *Prosiding Seminar Nasional Sanata Dharma Berbagi" Pengembangan, Penerapan Dan Pendidikan 'Sains Dan Teknologi' Pasca Pandemi"*. Yogyakarta: Sanata Dharma University Press, pp. 93, 2022, ISBN: 6236103968.