ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN DECISION TREE

Bayu wicaksono, Nuri Cahyono*
S1 Informatika, Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta
nuricahyono@amikom.ac.id

ABSTRAK

Instagram merupakan salah satu sosial media yang digunakan merepresentasikan diri, berinteraksi, dan mencari informasi. Kita dapat mengambil sekumpulan informasi ke dalam bentuk dataset untuk diolah lebih lanjut. Berkaitan dengan hal itu , Program Kampus Merdeka sebagai objek analisis, mengingat Program Kampus Merdeka adalah program pemerintah yang saat ini sedang dijalanan oleh Kemendikbud. Pengambilan *dataset* yang didapat dari kumpulan komentar Instragram, tool yang digunakan adalah phantombuster. Menggunakan bahasa pemrograman phyton dengan tools *Google Collab*, dengan Algoritma *Naïve Bayes Classifier* dan *Decision Tree* untuk membuat model sentimen. Hasil *scrapping* mendapatkan 1764 data , dan sesudah dilakukan *pre-processing* menjadi 1694 data. Dari sentimen analisis yang telah dilakukan diperoleh hasil dari penerapan Algortima *Complement Naïve Bayes* dan *Decision Tree*, sebelum dilakukan *SMOTE over-sampling*, perbandingan data positif dan negatif sebesar 35,06% banding 64,95% , dengan akurasi model *Decision Tree* 84% dengan skenario pembagian data 90:10 dan model *Complement Naïve Bayes* 81% pada skenario pembagian data 80:20. Setelah dilakukan balancing data menggunakan *SMOTE over-sampling*, akurasi pada model *Decision Tree* naik sebesar 1% dari 86% menjadi 85%, dengan skenario pembagian data 90:10, dan pada model *Complement Naïve Bayes* juga mengalami kenaikan sebesar 2%, dari 82% menjadi 83% dengan skenario pembagian data 80:20.

Kata kunci : *Sentimen Analisis, Naïve Bayes, Decision Tree,Machine Learning, Google Colab, phantombuster*

1. PENDAHULUAN

Program Kampus Merdeka adalah program pemerintah dari Kementerian Pendidikan, Kebudayaan,Riset,danTeknologi(Kemendikbudristek) yang memungkinkan mahasiswa mengikuti mata kuliah di luar kurikulum studi mereka selama 1 semester, sambil melaksanakan kegiatan di luar lingkungan perguruan tinggi selama 2 semester (Kampus Merdeka kemedikbud, 2022). Instagram adalah salah satu sosial media yang bisa digunakan merepresentasikan diri, berinteraksi, dan mencari informasi [4]. Kita dapat mengambil sekumpulan informasi ke dalam bentuk dataset untuk diolah lebih lanjut. Pengambilan *dataset* dibantu dengan tools yang bernama phantombuster.

Program Kampus Merdeka menuai berbagai tanggapan dari berbagai pihak, terutama dari kalangan mahasiswa dari platform Instagram. Kontroversi muncul karena pelaksanaan Program Kampus Merdeka oleh Kemendikbud dinilai belum sepenuhnya siap secara matang, dan juga terdapat hambatan terkait fasilitas uang saku [2].

NLP merupakan disiplin ilmu di mana individu berinteraksi dan beroperasi dengan kumpulan dokumen menggunakan berbagai alat analisis [3]. Analisis sentimen merupakan bagian dari Pemrosesan Bahasa Alami atau *Natural Language Processing* (NLP), yang melibatkan proses membantu mengenali konten dari suatu kumpulan data terkait suatu peristiwa, baik itu bersifat positif, negatif, atau netral [4].

Metode yang dapat digunakan dalam pembuatan model analisis sentiment, salah dua nya adalah menggunakan Algoritma *Decision Tree* dan *Complement Naïve Bayes*. *Decision Tree* memanfaatkan metode seleksi atribut untuk memilih atribut yang sedang diuji. Proses algoritma ini bertujuan untuk meningkatkan akurasi dengan mengeliminasi cabang-cabang pohon yang membawa (*noise*) ke dalam data [3]. *Naïve Bayes Classifier* adalah algoritma klasifikasi multikelas berbasis supervisi, yang mengaplikasikan teorema Bayes dengan mengasumsikan independensi bersyarat secara "*naive*" antara setiap pasangan variabel. *Naïve bayes* bekerja dengan cepat dan sederhana. *Complement Naïve Bayes* masuk ke dalam golongan *Naïve Bayes Classifier*. Penggunaan *Complement Naïve Bayes* (CNB) sangat cocok digunakan pada dataset yang tidak seimbang (*imbalance*) [5].

Dengan adanya hal tersebut maka menjadi pemikiran bagi penulis untuk melakukan sebuah sentiment analis pada platform Instagram terkait sebuah kasus yaitu "Analisis Sentimen Komentar Instagram pada Program Kampus Merdeka dengan Algortima *Complement Naive Bayes* dan *Decision Tree*". Berdasarkan informasi sebelumnya, masalah penelitian dapat dirumuskan dengan mengimplementasikan algoritma *Decision Tree* dan *Complement Naïve Bayes* untuk mengklasifikasikan sentimen atau opini dalam data Instagram menjadi kelas positif dan negatif. Selain itu agar masalah yang ditinjau lebih spesifik dan terarah, maka ruang lingkup pada masalah di batasi pada komentar yang berada di dalam komentar akun Instagram @kampusmerdeka.ri.

2. TINJAUAN PUSTAKA

Pada bagian ini memuat literatur ilmiah yang relevan dengan penelitian yang dilakukan saat ini.

2.1. Text Mining

Text Mining adalah suatu proses ekstraksi informasi dari sekelompok dokumen dengan menggunakan alat analisis tertentu. *Text mining* mampu memberikan solusi terkait pengolahan data, pengorganisasian, dan analisis teks tak terstruktur dalam jumlah besar [7]. *Text mining* juga merupakan proses semi otomatis dalam penggalian pola (pengetahuan dan informasi yang berguna) dari sejumlah data berukuran besar dan tidak terstruktur berupa teks. Teks akan diubah menjadi representasi numerik yang terstruktur [11].

2.2. Analisis Sentimen

Analisis Sentimen (*sentiment analysis*) termasuk ke dalam salah satu bidang *Natural Language Processing* (NLP) yang melibatkan proses membantu mengenali konten dari suatu kumpulan data terkait suatu peristiwa, baik itu bersifat positif, negatif, atau netral [4]. Analisis sentimen difokuskan ke dalam review klasifikasi berdasarkan polaritas. Analisis sentiment berdasarkan klasifikasinya dibagi menjadi 2 yaitu, klasifikasi pendapat atau fakta. Hal ini adalah proses yang sangat penting untuk mementukan sebuah dokumen yang memiliki opini dan dokumen yang menyimpulkan opini bernilai positif, negatif, atau netral [11].

2.3. Instagram

Media sosial ialah media yang dipakai pengguna dalam bentuk mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya. Indonesia menempati posisi keempat dunia atau jumlah pengguna Instagram tertinggi di Asia, yaitu sebanyak 63 juta pengguna aktif [12]. Salah satu platform media sosial yang populer khususnya di Indonesia yakni Instagram. Instagram ialah salah satu media sosial yang didirikan oleh Kevin Systrom dan Mike Krieger yang umumnya berfungsi dimana pengguna dapat memposting foto dan video dengan lampiran teks [6]. Kita dapat mengambil sekumpulan informasi ke dalam bentuk dataset untuk diolah lebih lanjut.

2.4. Program Kampus Merdeka

Program Kampus Merdeka adalah inisiatif pemerintah dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) yang memungkinkan mahasiswa mengikuti mata kuliah di luar kurikulum studi mereka selama 1 semester, sambil melaksanakan kegiatan di luar lingkungan perguruan tinggi selama 2 semester (Kampus Merdeka kemedikbud, 2022). Hal ini dilakukan untuk mengasah kemapuan mahasiswa untuk siap terjun ke dunia kerja. Pembelajaran di Kampus Merdeka merupakan kemandirian untuk mencari ilmu pengetahuan di lapangan, seperti inovasi, kemampuan, kepribadian, dan interkasi social. Berbagai kegiatan belajar yang ditawarkan dari Program Kampus Merdeka antara lain, Pertukaran pelajar, Magang, Studi Independen, dan Penelitian [3].

2.5. *Pre-processing*

Data *Pre-processing* disebut juga Normalisasi, diperlukan untuk mengatasi hal ini dengan tujuan untuk mengembalikan sebanyak mungkin bahasa pesan ke bahasa alami dengan menghilangkan ekspresi atipikal sehingga dapat meminimalkan *noise* atau data yang tidak konsisten . *Pre-processing* juga bertujuan mengubah data mentah menjadi data yang berkualitas yang nantinya akan diolah pada tahapan selanjutnya [7].

2.6. VADER Sentimen

Valence Aware Dictionary and Sentiment Reasoner (VADER) adalah metode yang digunakan sebagai model untuk analisis sentimen yang mampu menentukan data yang beragam melalui kektuan emosiaonal yang sesuai dengan kampus data *Lexicon* yang tersedia. Kamus *Lexicon* dapat menentukan tiga kategori sentimen, yaitu positif, negatif, dan netral. Keuntungan menggunakan VADER *polarity detection* adalah tersediannya kamus yang berisi nilai setiap kata. Hasil sentiment akan ditambahkan skor total (*Compound Score*). Kriteria pengelompokan nilai positif apabila *Compound Score* bernilai lebih dari 0, apabila terletak diantara -0 sampai dengan 0 maka kategori sentiment netral, dan yang terakhir apabila bernilai kurang dari 0 maka masuk ke dalam kategori negatif [11].

2.7. Google Collab

Google Collab merupakan sebuah *Intergrated Development Enviroment*(IDE) untuk pemrograman Phyton. Pemrosesan dilakukan menggunakn server Google yang memiliki performa yang tinggi. *Google Collab* menyediakan berbagai *library* yang banyak dibutuhkan [13].

2.8. TF-IDF

TF-IDF merupakan salah satu metode yang digunakan dalam pemrosesan Bahasa alami dan untuk meningkatkan akurasi dan relevansi hasil pencarian [16]. Dalam TF-IDF kata-kata yang tersusun dalam masih-masing komentar akan diberikan bobot dengan cara mengalikan nilai *Term Frequency* (TF) dengan *Inverse Document Frequency* (IDF). Mencari kemunculan masing-masing kata ke dalam *feature list* positif, negatif, dan netral. Komentar akan dibagi ke dalam data latih dan uji.

2.9. *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah alg (3)
klasifikasi multi kelas berbasis supervisi, yang mengaplikasikan teorema *Bayes* dengan mengasumsikan independensi bersyarat secara "*naive*"

antara setiap pasangan variabel. *Naïve Bayes* bekerja dengan cepat dan sederhana. *Complement Naïve Bayes* masuk ke dalam golongan *Naïve Bayes Classifier*. Penggunaan *Complement Naïve Baye*s (CNB) sangat cocok digunakan pada *dataset* yang tidak seimbang (*imbalance*) [5].

2.10. *Decision Tree*

Decision Tree memanfaatkan metode seleksi atribut untuk memilih atribut yang sedang diuji. Proses algoritma ini bertujuan untuk meningkatkan akurasi dengan mengeliminasi cabang-cabang pohon yang membawa (*noise*) ke dalam data [3].

2.11. *Synthetic Minority Over-sampling Technique* (SMOTE)

SMOTE adalah metode over-sampling, data pada kelas minoritas diperbanyak dengan cara replikasi pada kelas minoritas. Metode ini digunakan apabila jumlah data tidak mencukupi [14]. Algoritma yang bekerja pada SMOTE akan mengambil nilai selisih vector dari fitur kelas minoritas dan nilai *nearest neighbor* dari kelas minoritas, lalu mengalikan nilai tersebut dengan angka acak antara 0 sampai 1. Selanjutnya hasil tersebut ditambahkan dengan vector fiturnya sehingga mendapat nilai vector yang baru [15].

2.12. Evaluasi Model

Evaluasi model klasifikasi *Complement Naive Bayes* dan *Decision Tree* menggunakan *Confusion Matrix*. Dengan menggunakan *Confusion Matrix* kita dapat melihat mengevaluasi seberapa akurat model atau metode yang telah kita buat [4]. Melalui *Confusion Matrix* kita dapat menilai akurasi, presisi, *recall*, dan *f1-score*. Tabel *Confusion Matrix* dapat dilihat pada gambar dibawah.

Tabel 1. Confusion Matrix

	Nilai Nyata	
Nilai Prediksi	True Positif	False Negatif
	False Positif	True Negatif

Akurasi merupakan presentasi dari prediksi model yang telah dibuat. Nilai akurasi dapat dihitung pada persamaan.

$$akurasi = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \; x \; 100\% \qquad (1)$$

Presisi digunakan untuk menghitung berapa rasio atau keakuratan prediksi yang dilakukan oleh model.

$$precison = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (2)$$

Recall merupakan kemampuan classifier untuk menemukan semua kasus positif.

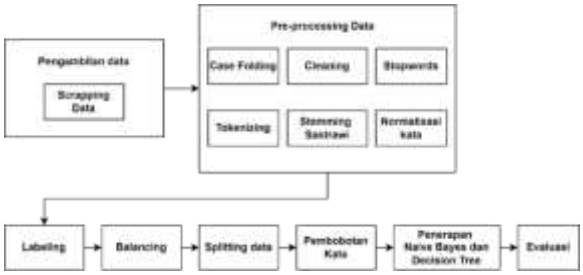
$$recall = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score merupakan rata-rata dari precision dan recall yang dibobotkan , dan memiliki presentase akurasi yang lebih rendah dari akurasi karena menanamkan *precision* dan *recall* ke dalam rumus perhitungan.

$$f1 - score = \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

3. METODE PENELITIAN

Penelitian ini melalui beberapa tahapan yang dapat diilustrasikan pada Gambar 1.



Gambar 1. Alur Penelitian

3.1. Pengambilan Data

Data yang diambil adalah hasil *scrapping* data pada platform Instagram menggunakan web phantombuster. Pengambilan data diambil dari akun resmi @kampusmerdeka.ri dari tanggal 26 Juli 2022 sampai 28 Juli 2023. Atribut yang dikumpulkan hanya diambil user, tanggal dan waktu, dan komentar. Pengambilan data dilakukan dengan cara memanfaatkan *link* dari setiap postingan dari akun @kampusmerdeka.ri, lalu akan di *copy-paste* ke dalam web phantombuster untuk di *generate*, nantinya akan menghasilkan beberapa komentar yang dapat di download. Penggunaan phantombuster sendiri sangat praktis, dimana di dalamnya terdapat banyak fitur yang dapat digunakan, khususnya untuk proses pengambilan data pada suatu *platform* tertentu, misal pada penelitian ini menggunakan Instagram. Terdapat banyak fitur diantaranya adalah fitur pengambilan komentar, pengambilan informasi profil, pengambilan informasi postingan, dan lain-lain.

3.2. *Preprocessing*

3.2.1. *Case folding*

Kalimat akan diubah menjadi huruf kecil semua (*lowercase*). Dalam analisis sentimen pada data komentar Instagram, langkah ini membantu mengurangi variasi dalam kata-kata dengan huruf kapital atau huruf kecil, memastikan konsistensi dalam analisis.

3.2.2. *Cleansing*

Penghapusan symbol, karakter, *hastag*(#), *username*(*@username*), *url*(*https://situs.com*), *email* (*nama@domain.com*) untuk mengurangi *noise* pada kalimat atau elemen lain yang tidak memberikan kontribusi pada analisis sentimen.

3.2.3. *Stopwords*

Stopwords merupakan pembuangan kata-kata umum yang sering muncul dalam teks dan biasanya tidak

memberikan kontribusi signifikan terhadap makna atau sentimen.

3.2.4. *Tokenizing*

Tokenizing merupakan pengubahan dari kalimat atau teks menjadi beberapa kata. Proses *tokenizing* penting dikarenakan dapat membantu mesin memahami bahasa manusia, sehingga lebih mudah untuk dianalisis.

3.2.5. *Stemming sastrawi*

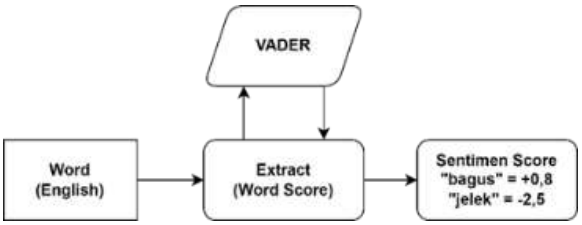
Penghapusan semua pengimbuhan awalan dan akhiran kata yang akan di ubah menjadi kata dasar. Penggunaan *library Sastrawi* dikarenakan kemampuan stemming dengan dukungan Bahasa Indonesia lebih baik dibandingkan dengan *library nltk*.

3.2.6. Normalisasi kata

Kata akan diubah ke dalam bentuk baku, memperbaiki kata gaul dan kata berbahasa inggris. Proses Normalisasi kata hampir sama dengan proses stemming, akan tetapi proses *stemming* hanya menghapus imbuhan pada awalan dan akhiran kata. Normalisasi kata membantu proses analisis sentimen untuk mendapatkan data yang lebih baik.

3.3. *Labeling*

Labeling dilakukan secara otomatis. Diawali dengan proses translate kata terlebih dahulu, dikarenakan library Vader Sentimen menggunakan kamus *lexicon* dengan berbahasa inggris. Setelah proses translate dilanjutkan pelabelan otomatis menggunakan library vader sentiment. Setiap kata dalam teks akan diberi skor sentimen berdasarkan kamus leksikal atau *lexicon*, dan skor setiap kata tersebut kemudian dijumlahkan untuk menghasilkan skor sentimen keseluruhan untuk teks tersebut. Setelah itu, Metode ini akan otomatis memberikan label, dengan parameter > 0 adalah label positif dan < 0 adalah label negatif. Proses penentuan polaritas kalimat didapatkan dari penyatuan atribut "compound" dari setiap kata yang tersedia. Keuntungan penggunaan VADER (*Valence Aware Dictionary and Sentiment Reasoner*) adalah menggabungkan analisis kualitatif dan validasi empiris menggunakan kebijaksanaan dan penilaian manusia [11].



Gambar 2. Alur Vader Sentimen

3.4. *Balancing Data*

Data yang digunakan merupakan data *imbalance* atau data tidak seimbang. Kondisi tersebut apabila pada suatu *dataset* memiliki data yang sangat besar pada kelas mayoritas dibandingkan dengan kelas minoritas. Perbedaan tersebut menyebabkan model klasifikasi tidak dapat memperediksi dengan tepat. Untuk mengatasi permasalahan tersebut diperlukannya balancing data atau penyeimbangan data. Salah satu teknik yang sering digunakan yaitu *Syntetic Minority Over-sampling Technique* (SMOTE) metode *Oversampling* [14].

3.5. Pembobotan kata

Pembobotan kata dilakukan menggunakan *library* TF-IDF vectorizer. Dalam tahapan ini komentar akan diberikan bobot dengan cara mengalikan nilai *Term Frequency* (TF) dengan *Inverse Document Frequency* (IDF). Secara sederhana metode TF-IDF digunakan untuk mengetahui kata yang sering muncul di dalam dataset atau dokumen. Setelah dilakukan pembobotan kata, selanjutnya akan dilakukan pada tahap *splitting data*.

3.6. *Splitting data*

Splitting data dilakukan dengan membagi tiga kategori pembagian data, yaitu dengan perbandingan (90:10, 80:20, dan 70:30). Data *training* yaitu data yang akan di olah pada metode yang digunakan, yang hasilnya adalah sebagai bahan prediksi terhadap data testing. Sedangkan data *testing* adalah data yang akan diuji dan di prediksi. Dengan tiga skenario splitting data, akan dilihat model dengan nilai akurasi terbaik pada skenario pembagian data yang mana. Setelah dilakukan *splitting data*, selanjutnya akan dilakukan pada tahap modeling atau implementasi model *Complement Naïve Bayes* dan *Decision Tree*.

3.7. Penerapan *Complement Naïve Bayes* dan *Decision Tree*

Pembuatan model menggunakan menerapkan 2 algortima yang akan dibandingkan, yaitu algortima *Complement Naïve Bayes* (CNB) dan *Decision Tree*. Penggunaan algortima *Complement Naïve Bayes* (CNB) dikarenakan model akan diimplementasikan pada dataset yang tidak seimbang (*imbalance*). Dikarenakan *Complement Naïve Bayes cocok* untuk dataset yang tidak seimbang. Selain itu nantinya model *Complememnt Naïve Bayes* juga akan diuji pada data yang sudah seimbang (*balance*). Pada proses ini memberikan hasil klasifikasi, yaitu sentiment negatif dan positif pada model yang telah dibuat.

3.8. Evaluasi model

Evaluasi hasil model menggunakan *Confusion Matrix*. Di dalamnya berisi akurasi, presisi, *recall*, dan *f1-score* hasil dari proses hasil klasifikasi. Akurasi merupakan presentasi dari sebuah prediksi model yang dibuat. Presisi adalah tingkat akurasi antara data yang diterima pemakai. *Recall* adalah kualitas hasil yang

relevan dengan yang ditampilkan oleh hasil prediksi. *F1-Score* adalah perhitungan evaluasi yang menggabungkan presisi dan *recall*. Dengan menggunakan *Confusion Matrix* diharapkan dapat mengetahui seberapa akurat metode yang dievaluasi.

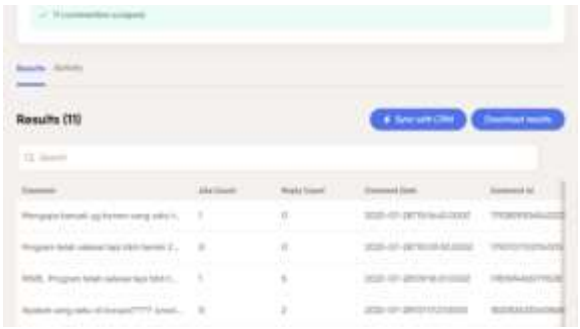
4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan bahasa pemrograman python dengan *tools* google colaboratoy dan mengimplementasikan algoritma *Complement Naïve Bayes* dan *Decision Tree* untuk membuat model analisis sentiment terhadap komentar Instagram pada Program Kampus Merdeka.

4.1. *Scrapping Data*

Scrapping data diambil pada platform Instagram melalui sebuah akun @kampusmerdeka.ri. Scrapping data Instagram diambil dalam bentuk teks. Pengambilan dataset dibantu dengan web phantombuster, dengan memanfaatkan link tautan Instagram. Data diambil dari tanggal 26 Juli 2022 sampai 28 Juli 2023 berjumlah 1764 komentar. Atribut yang dikumpulkan hanya diambil user, tanggal dan waktu, dan komentar.

Pengambilan data dilakukan dengan cara memanfaatkan *link* dari setiap postingan dari akun @kampusmerdeka.ri, lalu akan di *copy-paste* ke dalam web phantombuster untuk di *generate*, nantinya akan menghasilkan beberapa komentar yang dapat di *download*. Penggunaan phantombuster sendiri sangat praktis, dimana di dalamnya terdapat banyak fitur yang dapat digunakan, khususnya untuk proses pengambilan data pada suatu platform tertentu, misal pada penelitian ini menggunakan Instagram. Terdapat banyak fitur diantaranya adalah fitur pengambilan komentar, pengambilan informasi profil, pengambilan informasi postingan, dan lain-lain. Hasil sampel pengambilan komentar pada web phantombuster dapat dilihat pada gambar 3. dibawah.



Gambar 3. Hasil Scrapping data phantombuster

4.2. *Preprocessing Data*

Proses *preprocessing* sentimen dari data komentar yang diperoleh dari web scrapping yaitu phantombuster. Melibatkan langkah-langkah penting seperti case folding mengubah kalimat menjadi huruf kecil semua(*lowercase*) untuk menjaga konsistensi dalam proses analisis, cleansing untuk pembersihan data dari elemen yang tidak relevan (misal: emoji, hastag, email, karakter), penghapusan stopwords, tokenisasi untuk membagi teks menjadi kata-kata individual, *stemming* atau *lemmatization* untuk mengubah kata-kata menjadi bentuk dasar, normalisasi kata untuk mengubah ke bentuk baku, mengubah ejaan berbahasa inggris. Setelah proses preprocessing, data kemudian dapat dilabeli dengan menentukan sentimen dari komentar-komentar tersebut seperti, positif maupun negatif, untuk digunakan sebagai data latih dalam analisis sentimen menggunakan metode seperti *Complement Naive Bayes* dan *Decision Tree*. Tujuan dari proses ini adalah untuk membersihkan, mengurangi dimensi, dan mengonversi data teks menjadi format yang dapat diproses oleh algoritma, sehingga siap untuk diolah lebih lanjut.

4.2.1. *Case Folding*

Pada tahap ini semua kata dalam komentar diubah menjadi huruf kecil (*lowercase*). Dalam analisis sentimen pada data komentar Instagram, langkah ini membantu mengurangi variasi dalam kata-kata dengan huruf kapital atau huruf kecil, memastikan konsistensi dalam analisis.

Tabel 2. *Case Folding*

Sebelum	Sesudah
Apakah uang saku di korupsi???? @nadiemmakarim	apakah uang saku di korupsi???? @nadiemmakarim
Kak untuk mahasiswa Universitas Terbuka apakah bisa mengikuti program magang? Karena dari kampus nggak ada kegiatan magang @kampusmerdeka.ri @nadiemmakarim	kak untuk mahasiswa universitas terbuka apakah bisa mengikuti program magang? karena dari kampus nggak ada kegiatan magang @kampusmerdeka.ri @nadiemmakarim

4.2.2. *Cleansing*

Penghapusan symbol, karakter, *hastag*(#), *username(@username*), *url(https://situs.com*), *email (nama@domain.com*) untuk mengurangi *noise* pada kalimat atau elemen lain yang tidak memberikan kontribusi pada analisis sentimen.

Tabel 3. Cleansing

Sebelum	Sesudah
Apakah uang saku di korupsi???? @nadiemmakarim	apakah uang saku dikorupsi
kak untuk mahasiswa universitas terbuka apakah bisa mengikuti program magang? karena dari kampus nggak ada kegiatan magang @kampusmerdeka.ri @nadiemmakarim	kak untuk mahasiswa universitas terbuka apakah bisa mengikuti program magang karena dari kampus nggak ada kegiatan magang ri

4.2.3. *Stopwords*

Stopword adalah kata-kata umum yang sering muncul dalam teks dan biasanya tidak memberikan kontribusi signifikan terhadap makna atau sentimen. Dalam konteks analisis sentimen pada data komentar Instagram, contoh *stopwords* meliputi kata-kata seperti "dan", "atau", dan lainnya. Menghapus Stopwords dari data membantu fokus pada kata-kata yang lebih bermakna dalam analisis sentimen, karena kata-kata ini tidak membawa nilai sentimen yang signifikan dan dapat mengganggu proses analisis.

Tabel 4. Stopwords

Sebelum	Sesudah
apakah uang saku dikorupsi	uang saku korupsi
kak untuk mahasiswa universitas terbuka apakah bisa mengikuti program magang karena dari kampus nggak ada kegiatan magang ri	kak mahasiswa universitas terbuka mengikuti program magang kampus nggak kegiatan magang ri

4.2.4. *Tokenizing*

Tokenizing adalah proses mengubah teks menjadi unit-unit yang lebih kecil, yang disebut token. Dalam konteks analisis sentimen pada data komentar Instagram penelitian ini, tokenizing melibatkan pemisahan teks menjadi kata-kata individual. Proses *Tokening* penting dikarenakan dapat membantu mesin memahami Bahasa manusia, sehingga lebih mudah untuk dianalisis. Misalnya, kalimat "komen uang saku dibayar emang perjanjian tau pendaftaran individu bayar terimakasih" akan dipecah menjadi token-token yang terdiri dari “komen”, “uang”, “saku”, “dibayar”, “emang”,“perjanjian”, “tau”,pendaftaran”, “individu”, “bayar”, “terimakasih”.

Tabel 5. *Tokenizing*

Sebelum	Sesudah
uang saku korupsi	{”uang”, ”saku”, ”korupsi”}
kak mahasiswa universitas terbuka mengikuti program magang kampus nggak kegiatan magang ri	{”kak”,”mahasiswa”,”universitas”, ”terbuka”,”mengikuti”,”program”, ”magang”,”kampus”,”nggak”, ”kegiatan”, ”magang”, ”ri”

4.2.5. *Stemming Sastrawi*

Stemming merupakan proses dalam pemrosesan teks yang bertujuan untuk mengubah kata-kata menjadi bentuk dasarnya dengan menghapus awalan dan akhiran kata. *Stemming Sastrawi* menggunakan algoritma bahasa Indonesia untuk menghasilkan kata dasar dari kata-kata yang diinfleksikan atau diubah bentuknya. Penggunaan *library Sastrawi* dikarenakan kemampuan stemming dengan dukungan Bahasa Indonesia lebih baik dibandingkan dengan *library nltk*. Misalnya, kata "berlari", "berlaku", dan "berlalu" akan diubah menjadi kata dasar "lari", "laku", dan "lalu". *Stemming* sangat penting dalam *preprocessing* data teks untuk mengurangi variasi kata dan meningkatkan konsistensi dalam analisis.

Tabel 6. *Stemming Sastrawi*

Sebelum	Sesudah
uang saku korupsi	uang saku korupsi
kak mahasiswa universitas terbuka mengikuti program magang kampus nggak kegiatan magang ri	kak mahasiswa universitas buka ikut program magang kampus nggak giat magang ri

4.2.6. Normalisasi Kata

Normalisasi kata adalah tahap dalam *preprocessing teks* di mana kata akan diubah menjadi bentuk baku, mengubah ejaan berbahasa inggris dan memperbaiki kata gaul. Proses Normalisasi kata hampir sama dengan proses *stemming*, akan tetapi proses *stemming* hanya menghapus imbuhan pada awalan dan akhiran kata. Normalisasi kata membantu proses analisis sentimen untuk mendapatkan data yang lebih baik.

Tabel 7. Normalisasi Kata

Sebelum	Sesudah
kak mahasiswa universitas buka ikut program magang kampus nggak giat magang ri	kak mahasiswa universitas buka ikut program magang kampus tidak giat magang ri
cair magang selesai hubung helpdesk beberapakali skrg kembang mohon bantuanya moga team msib mudah rizkinya lancar urus ri	cair magang selesai hubung helpdesk beberapakali sekarang kembang mohon bantuan semoga team msib mudah rizkinya lancar urus ri

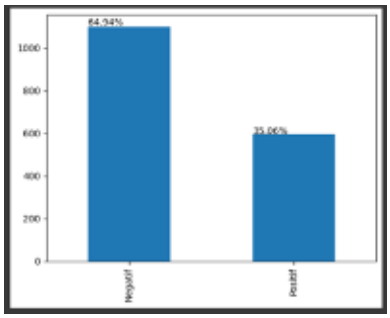
4.3. *Labeling*

Labeling pada penelitian ini menggunakan metode VADER (*Valence Aware Dictionary and Sentiment Reasoner*). Dimana proses memberikan label sentimen (misalnya, positif, negatif, atau netral) kepada teks berdasarkan analisis sentimen menggunakan metode VADER. Metode ini menggunakan kamus leksikal atau *lexicon* yang sudah ada dan aturan heuristik untuk menentukan sentimen dari kata-kata dalam teks dalam berbahasa inggris. Oleh karena itu sebelum penerapan Metode ini, dilakukan proses translate kalimat terlebih dahulu. Setiap kata dalam teks akan diberi skor sentimen berdasarkan kamus leksikal atau lexicon, dan skor setiap kata tersebut kemudian dijumlahkan untuk menghasilkan skor sentimen keseluruhan untuk teks tersebut. Setelah itu, Metode ini akan otomatis memberikan label, dengan parameter > 0 adalah label positif dan < 0 adalah label negatif. Keuntungan penggunaan VADER (*Valence Aware Dictionary and Sentiment Reasoner*) adalah menggabungkan analisis kualitatif dan validasi empiris menggunakan kebijaksanaan dan penilaian manusia.

Tabel 8. *Labeling*

Text	Compund Score	Label
bismillah i hope you pass all my friends amen	0.7184	Positif
failed try msib wait to fight the spirit that failed	-0.8176	Negatif

Berdasarkan skor tersebut, teks akan diberi label sentimen yang sesuai. Proses ini membantu secara otomatis menganalisis sentimen teks dengan cepat dan efisien tanpa memerlukan label manual dari manusia, sehingga cocok untuk analisis besar-besaran pada data teks yang diperoleh dari sumber seperti media sosial, misalnya Instagram.

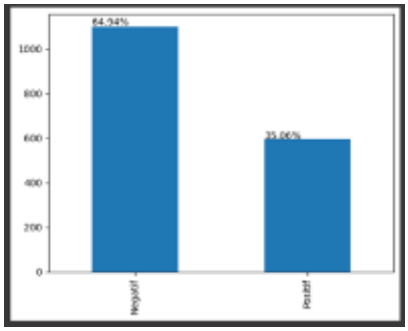


Gambar 4. Visualiasi persebaran data

Visualisasi data pada gambar 4. ini menunjukkan persebaran data sentimen positif sebesar 65.17% dan negatif sebesar 34.83%. Kondisi ini menunjukkan ketidakseimbangan dalam dataset, dengan sentimen positif lebih dominan. Untuk mengatasi ketidakseimbangan ini, penelitian akan menerapkan teknik balancing data.

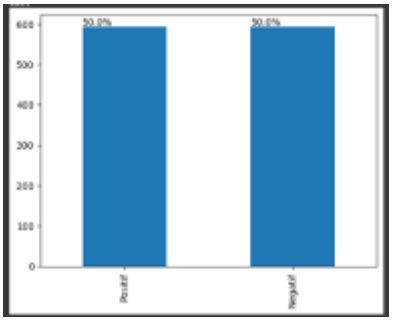
4.4. *Balancing Data*

Balancing data dilakukan dengan menerapkan *SMOTE- oversampling*. *Balancing data* dilakukan dengan meningkatkan kelas minor (pada penelitian ini kelas minor adalah kelas positif) dengan membangkitkan data baru (data sintetis) agar data seimbang. Sebelum dilakukan *Balancing data*, persebaran data negatif sebesar 64,94% dengan populasi persebaran data negatif 1100 data, dan persebaran data positif sebesar 35,06% dengan persebaran data positif sebesar 594 data. Persebaran data positif dan negatif dapat dilihat pada gambar 5. dibawah.



Gambar 5. Visualisasi *Imbalance data*

Setelah dilakukan Balancing data, terlihat nilai negatif dan positif sebesar 50% berbanding dengan 50%. Persebaran data positif dan negatif setelah dilakukan balancing data dengan menggunakan SMOTE-oversampling dapat dilihat pada gambar 6. dibawah.



Gambar 6. Visualiasi setelah *balancing data*

4.5. Pembobotan Kata

TF-IDF (*Term Frequency-Inverse Document Frequency)* melibatkan *tokenization* dan pembentukan dokumen, di mana setiap kata dalam setiap dokumen diberi bobot berdasarkan frekuensi kemunculannya (*Term Frequency*). Selanjutnya, dihitung bobot *inverse document frequency* (IDF) untuk menentukan seberapa penting suatu kata dalam seluruh dataset. Bobot kata akhir (TF-IDF) diperoleh dengan mengalikan bobot *Term Frequency* (TF) dengan bobot *Inverse Document Frequency* (IDF). TF-IDF adalah pemoboban kata yang umum dan sering digunakan dalam proses analisis sentimen dengan hasil yang akurat. Hasilnya adalah vektor numerik yang merepresentasikan teks, memungkinkan model machine learning untuk memproses dan menganalisis informasi sentimen dengan lebih baik. Hasil TF-IDF dapat dilihat pada tabel 9. dibawah.

Tabel 9. Frekuensi kemunculan kata

Kata	TF-IDF
min	95.053724
please	47.969335
email	44.869503
bismillah	43.818670
long	43.807991
test	41.200007
server	41.095149
internship	37.118702
diversity	31.664450
register	31.615734
pass	28.914791
website	27.693842
independent	0.015468
thank	0.014864
msib	0.014011

4.6. *Spliting Data*

Spliting Data pada penelitian ini menggunakan pembagian data training dan data testing sebanyak tiga variasi rasio pembagian data, diantaranya (90:10, 80:20, dan 70:30). Pembagian data ini akan

diterapkan pada dua tipe data, yaitu *imbalance data* dan *balance data*. Skenario pembagian data ini untuk melihat nilai akurasi model yang paling baik berada pada skenario pembagian data yang mana.

Tabel 10. Skala pembagian data

Data Training	Data Testing
90	30
80	20
70	30

4.7. Penerapan *Complement Naïve Bayes* dan *Decision Tree*

Dalam penelitian ini, penerapan *Naïve Bayes* dan *Decision Tree* pada sentimen analisis menitikberatkan pada akurasi, presisi, *recall*, dan *F1-score*. Langkah-langkah utama melibatkan pembagian data dengan variasi rasio pembagian data (90:10, 80:20, 70:30), pelatihan model, dan evaluasi performa dengan metrik yang ditekankan. Akurasi memberikan gambaran umum tentang keberhasilan model, presisi menilai seberapa banyak prediksi positif yang benar, recall menunjukkan sejauh mana model mengidentifikasi keseluruhan kelas positif, dan *F1-score* menggabungkan presisi dan *recall* untuk memberikan ukuran holistik terhadap performa model.

Penelitian ini juga melibatkan uji dengan balancing data dan imbalance data untuk memahami pengaruhnya terhadap metrik-metrik tersebut, memberikan wawasan yang komprehensif terkait pemilihan model terbaik dalam konteks analisis sentimen. Pemilihan metode *Complement Naïve Bayes* dikarenakan pada penelitian ini menggunakan data yang tidak *balance*. *Complement Naïve Bayes* cocok pada *dataset* yang tidak seimbang (*imbalance*). Meskipun pada penelitian ini nantinya akan dilakukan *balancing data*, namun peneliti ingin membandingan model, apakah pada data balance, hasil akurasi dari model *Complement Naïve Bayes* akan mengalami penurunan atau kenaikan.

4.7.1. *Complement Naïve Bayes*

Pada pembagian data 80:20, penerapan model dalam imbalance dataset mencapai presisi sebesar 80%, *recall* sebesar 78%, *F1-score* sebesar 79%, dan akurasi sebesar 81%. Sedangkan pada proporsi pembagian data 90:10 dan 70:30, terdapat sedikit variasi dalam metrik evaluasi performa model, namun secara keseluruhan, model menunjukkan kinerja yang relatif stabil dengan nilai presisi, *recall*, *F1-score*, dan akurasi yang cukup tinggi. Hasil model CNB pada *imbalance dataset* dapat dilihat pada table 10. dibawah.

Tabel 11. *imbalance dataset* CNB

Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	79%	79%	79%	81%
80:20	80%	78%	79%	81%
70:30	79%	77%	78%	80%

Kemudian, berdasarkan hasil dengan *balance dataset* yang mencatat performa model dengan variasi pembagian data (90:10, 80:20, dan 70:30), dapat disimpulkan bahwa dapat diamati bahwa untuk semua proporsi pembagian data, model menunjukkan kinerja yang konsisten dengan nilai presisi, *recall*, *F1-score*, dan akurasi yang tinggi. Hal ini menunjukkan bahwa model *Complement Naïve Bayes* dapat secara konsisten menghasilkan prediksi yang akurat dan konsisten pada dataset yang seimbang. Hasil model CNB pada *balance dataset* dapat dilihat pada table 12. dibawah.

Tabel 12. *balance dataset* CNB

Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	81%	81%	81%	81%
80:20	83%	83%	83%	83%
70:30	81%	81%	81%	81%

4.7.2. *Decision Tree*

Pemodelan *Decision Tree* pada keadaan *dataset* tidak seimbang (*imbalance*) dapat diamati bahwa untuk semua proporsi pembagian data, model Decision Tree menunjukkan kinerja yang relatif konsisten dengan nilai presisi, *recall*, *F1-score*, dan akurasi yang cukup tinggi. Namun, perlu dicatat bahwa meskipun nilai akurasi tinggi, nilai *recall* tampaknya sedikit lebih rendah dibandingkan dengan presisi dan akurasi. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mengidentifikasi *instance* negatif daripada *instance* positif. Hasil Pemodelan dengan *imbalance dataset* dapat dilihat pada tabel 13. dibawah.

Tabel 13. *imbalance dataset Decision Tree*

Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	83%	80%	81%	84%
80:20	83%	80%	81%	83%
70:30	82%	80%	81%	83%

Sedangkan, akurasi model dengan *balance dataset* menggambarkan performa model pada kondisi *dataset* yang seimbang. Dapat diamati bahwa untuk semua proporsi pembagian data, model *Decision Tree* menunjukkan kinerja yang relatif konsisten dengan nilai presisi, *recall*, *F1-score*, dan akurasi yang cukup tinggi. Khususnya, pada proporsi pembagian data 90:10, model mencapai presisi, *recall*, *F1-score*, dan akurasi sebesar 85%, menunjukkan kinerja yang optimal. Hasil Pemodelan dengan *balance dataset* dapat dilihat pada tabe 14. dibawah.

Tabel 14. akurasi model dengan balance dataset

Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	85%	85%	85%	85%
80:20	84%	84%	83%	83%
70:30	81%	81%	81%	81%

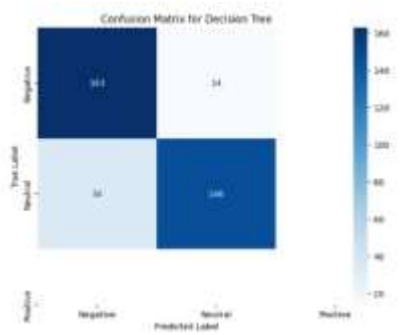
4.8. Evaluasi Hasil

Hasil dari penerapan Algortima *Complement Naïve Bayes* dan *Decision Tree*, sebelum dilakukan balancing data hasil akurasi *Complement Naïve Bayes* didapatkan akurasi terbaik sebesar 81% dengan skenario pembagian data 80:20 dan *Decision Tree* sebesar 84% dengan skenario pembagian data 90:10 pada *dataset* yang tidak seimbang (*imbalance*). Setelah dilakukan balancing data, akurasi Model *Complement Naïve Bayes* mengalami kenaikan sebesar 2%, dari 81% menjadi 83% dengan skenario pembagian data 80:20 dan pada model *Decision Tree* juga mengalami kenaikan akurasi sebesar 1%, dari 84% menjadi 85% dengan pembagian data 90:10. Pada penelitian ini juga ditambahkan Cabang Naïve Bayes lain, yaitu *Multinomial Naïve Bayes* (MNB) untuk membandingkan lebih baik mana antara model *Complement Naïve Bayes* atau *Multinomial Naïve Bayes* pada kondisi *dataset* yang seimbang atau balance.

Tabel 15. Perbandingan Model Naïve Bayes dan Decision Tree Confusion Matrix

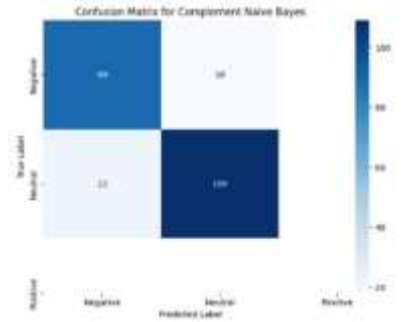
Imbalance Dataset				
Decision Tree				
Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	83%	80%	81%	84%
Complement Naïve Bayes				
80:20	80%	78%	79%	81%
Balance Dataset				
Decision Tree				
Splitting Data	Presisi	Recall	F1-Score	Akurasi
90:10	85%	85%	85%	85%
Complement Naïve Bayes				
80:20	83%	83%	83%	83%
Multinomial Naïve Bayes				
80:20	82%	82%	82%	82%

Dari *Confusion Matrix* diatas, Pada *dataset* yang tidak seimbang (*imbalance dataset*), terlihat bahwa performa model *Decision Tree* dan *Complement Naïve Bayes* (CNB) cenderung lebih rendah dibandingkan dengan performa pada *dataset* yang seimbang (*balance*). Hal ini terlihat dari penurunan nilai presisi, recall, F1-score, dan akurasi pada model *Decision Tree* dan *Complement Naïve Bayes* (CNB) pada *dataset* tidak seimbang (*imbalance*). Pada *dataset* yang seimbang, terlihat bahwa model *Decision Tree* memiliki performa yang lebih baik daripada model *Complement Naïve Bayes* (CNB). Hal ini terlihat dari nilai presisi, *recall*, *F1-score*, dan akurasi yang lebih tinggi pada model *Decision Tree* dibandingkan dengan model *Complement Naïve Bayes* (CNB) pada proporsi pembagian data 90:10 dan 80:20. Pada Model *Multinomial Naïve Bayes*, dari segi Presisi, *Recall, F1-Score* dan Akurasi, memiliki nilai yang lebih rendah daripada model *Complement Naïve Bayes* (CNB).



Gambar 7. *Confusion Matrix Decision Tree balance dataset*

Pada gambar 7. Confusion matriks menunjukkan ketepatan model dalam mengidentifikasi ulasan yang benar-benar negatif (*True Negatives*) sebanyak 163 kasus, dan positif (*True Positives*) sebanyak 146 kasus. Meskipun demikian, terdapat beberapa kesalahan dalam klasifikasi, di mana model memberikan prediksi negatif palsu (*False Negatives*) sebanyak 34 kasus dan positif palsu (*False Positives*) sebanyak 14 kasus.



Gambar 8. *Confusion Matrix Complement Naïve Bayes balance dataset*

Pada gambar 8. *Confusion matriks* menunjukkan ketepatan model dalam mengidentifikasi ulasan yang benar-benar negatif (*True Negatives*) sebanyak 89 kasus, dan positif (*True Positives*) sebanyak 109 kasus. Meskipun demikian, terdapat beberapa kesalahan dalam klasifikasi, di mana model memberikan prediksi negatif palsu (*False Negatives*) sebanyak 22 kasus dan positif palsu (*False Positives*) sebanyak 18 kasus.

5. KESIMPULAN DAN SARAN

Berdasarkan Peneelitian yang telah dilakukan tentang " Analisis Sentimen Komentar Instagram Pada Program Kampus Merdeka Dengan Algortima *Complement Naive Bayes* dan *Decision Tree*", dengan menggunakan 1764 data sebelum dilakukan *preprocessing*, dan 1694 data setelah dilakukan *preprocessing*. *Labeling* dilakukan dengan cara otomatis menggunakan metode VADER (*Valence Aware Dictionary and Sentiment Reasoner*), menggunakan 2 kategori Sentimen yaitu Positif dan Negatif. Pada Penelitian ini menggunakan data yang tidak seimbang atau *imbalance* data, maka dilakukan *Balancing* data menggunakan *SMOTE over-sampling*. Sebelum dilakukan *SMOTE over-sampling*,

perbandingan data positif dan negatif sebesar 35,06% banding 64,95% , dengan akurasi model *Decision Tree* 84% dengan skenario pembagian data 90:10 dan model *Complement Naïve Bayes* 81% pada skenario pembagian data 80:20. Setelah dilakukan balancing data menggunakan *SMOTE over-sampling*, akurasi pada model *Decision Tree* naik sebesar 1% dari 84% menjadi 85%, dengan skenario pembagian data 90:10, dan pada model *Complement Naïve Bayes* juga mengalami kenaikan sebesar 2%, dari 81% menjadi 83% dengan skenario pembagian data 80:20.

Diharapkan pada penelitian selanjutnya pemberian *labeling* atau sentimen dilakukan dengan cara manual di bawah pengawasan para ahli bahasa, guna untuk membandingkan hasil sentimen, Penelitian selanjutntya dapat melakukan percobaan dengan 3 kategori sentimen (Positif, Negatif, dan Netral), dan penambahan Metode yang berbeda untuk menemukan model klasifikasi yang lebih baik.

DAFTAR PUSTAKA

[1] Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. "Program Kampus Merdeka." [Online]. Available: https://kampusmerdeka.kemdikbud.go.id/program. [Accessed: Feb. 7, 2024].

[2] Vantissha. D, Fauziah. Y, "Pengaruh Implementasi Kegiatan Merdeka Belajar Kampus Merdeka (MBKM) Terhadap Mahasiswa di Porgram Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Esa Unggul", Vol.08, No.02, Desember 2021.

[3] Arsi. P, Berlilana, and Rokhman.K.A, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online", Joism : Jurnal Of Information System Management, Vol.02, No.02, 2715-3088, 2021.*and Information 2015 (ICESTI 2015)* (pp. 337-343). Springer Singapore.

[4] A. Oktavia Praneswara and N. Cahyono, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," Indonesian Journal of Computer Science Attribution, vol. 12, no. 6, p. 3925.

[5] Zami.F.A, Udayanti.E.D, Pratiwi.Y.A, and Sani.R.R, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia", Vol.13, No.02, Jurnal Masyarakat Informatika, 2086-4930, 2022.

[6] Kurniawan.H, Srtika.M, and Anisah.N, "Penggunaan Media Sosial Instagram Dalam Meningkatkan Literasi Kesehatan Pada Mahasiswa", Vol.04, No.02, 2021.

[7] I. Habib Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, 2023.

[8] Hermawan.A, Wibowo.A.P, and Setiawati.I "Implementasi Decision Tree Untuk Mendiagnosis Penyakit Liver", Joism : Jurnal Of Information System Management, Vol.01, No.01, 2019.

[9] Widowati.T.T, Sadikin.M,"Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine", Vol.11, No.02, 2549-3108, November 2020.

[10] Angdresey.A, Wongkar.M, "Sentiment Analysis Using Naive Bayes Algorithm of the Data Crawler: Twitter", Univ of Science and Technology, 2020.

[11] Abimanyu.D, Budianita.E, Cynthia.E.P, Yanto.F, Yusra, "Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER", Jurnal Nasional Komputasi dan Teknologi Informasi, 2621-3052, Vol. 5 No. 3, Juni 2022.

[12] Anisah.N, Sartika.M, Kurniawan.H, "Penggunaan Media Sosial Instagram Dalam Meningkatkan Literasi Kesehatan Pada Mahasiswa", : 2598-6031, Vol. No. Tahun 2021.

[13] Guntara.R.G, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7", Jurnal Teknologi Dan Sistem Informasi Bisnis, Hal. 55-60, Vol. 5 No. 1 Januari 2023.

[14] Sutoyo.E, Fadlurrahman.M.Asri, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network", 2548-9364, Vol. 6 No. 3 Desember 2020.

[15] Astuti.F.D, Lenti.F.N, "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN", Jurnal JUPITER, Hal. 89 -98, Vol. 13 No. 1 Bulan April Tahun 2021.

[16] Isabela.I, Septiani.D, "Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks", SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia, Vol. 01, No. 2, 2807-9108, Maret 2022.