

Analisis Sentimen Terhadap Program Kampus Merdeka Pada Twitter Menggunakan Metode Naïve Bayes, Union dan *Synthetic Minority Over Sampling Technique* (SMOTE)

Jefina Tri Kumalasari¹, Agustiena Merdekawati²

¹Universitas Bina Sarana Informatika, jefina.jtk@bsi.ac.id, Jakarta, Indonesia

²Universitas Bina Sarana Informatika, agustiena.atd@bsi.ac.id, Jakarta, Indonesia

Informasi Makalah

Submit : Oktober 20, 2022
Revisi : November 11, 2022
Diterima : Juni 1, 2023

Kata Kunci :

Analisa Sentimen
Kampus merdeka
Naïve bayes
UNION
SMOTE

Abstrak

Program Kampus Merdeka bertujuan menyiapkan mahasiswa menghadapi tantangan perubahan sosial, dan kemajuan teknologi yang semakin cepat. Kampus Merdeka memberikan peluang mahasiswa untuk mengambil materi program atau jurusan lain atau yang sama pada kampus yang lain sehingga diharapkan dapat menghasilkan lulusan yang dapat menghadapi dinamika masyarakat pada lingkungan usaha dan industri. Twitter salah satu media yang digemari digunakan sebagai media bersosialisasi, bertukar informasi dan berpendapat sehingga sesuai sebagai sumber analisa sentiment. Mengumpulkan, mengenali, mengekstrak opini dalam bentuk teks sehingga dapat digunakan sebagai informasi yang bermanfaat. Algoritma klasifikasi Naïve Bayes memprediksi apakah suatu opini bernilai positif atau negative. Menganalisa data dengan jumlah besar membutuhkan operator Union. Variable yang digunakan dalam penelitian ini yaitu hasil tweet dengan kata kunci kampus merdeka dan hasil sentiment sebagai label. Data yang diambil berjumlah 680, sedangkan data yang tidak diberi label berjumlah 430 yang merupakan data raw yang berasal dari media social Twitter berisi ungkapan, pertanyaan, pengalaman pengguna. Berdasarkan hasil analisa akumulasi sentiment menghasilkan nilai positif 86% dibandingkan yang memberikan nilai negative 14%. Sehingga dapat disimpulkan Program Merdeka Belajar memberikan opini baik. Selain itu untuk mengoptimalkan data yang tidak seimbang digunakan Synthetic Minority Oversampling Technique (SMOTE) menghasilkan keakuratan data sebesar 99.21% sedangkan tanpa SMOTE keakuratan data hanya 77.60%. Sehingga dapat disimpulkan Program Merdeka Belajar memberikan opini baik.

Abstract

The Merdeka Campus Program aims to prepare students to face the challenges of social development and increasingly rapid technological advances. With the program, it is expected to be

Jefina Tri Kumalasari
Email: jefina.jtk@bsi.ac.id

able to answer the challenges of higher education to produce graduates who are in accordance with the demands of the business and industrial world, as well as the dynamics of the society of the business and industrial world. Twitter is one of the media used for socializing, and exchanging information and opinions so it is suitable as a source of sentiment analysis. Collect, identify, and extract opinions in the form of the text so that they can be used as useful information. The Naïve Bayes classification algorithm predicts whether an opinion is positive or negative. Analyzing large amounts of data requires the Union operator. The variables used in this study are the results of tweets with the keyword independent campus and sentiment results as labels. The data taken was 680, while the data that was not labeled was 430 which was raw data originating from social media Twitter containing expressions, questions, and user experiences. Based on the accumulated sentiment analysis results produces a positive value of 86% compared to a negative value of 14%. So it can be concluded that the Independent Learning Program gives a good opinion. In addition to optimizing unbalanced data, the Synthetic Minority Oversampling Technique (SMOTE) is used, resulting in 99.21% data accuracy, while without SMOTE the data accuracy is only 77.60%. So it can be concluded that the Independent Learning Program gives a good opinion.

1. Pendahuluan

Pemuda sebagai pondasi dalam pembangunan Indonesia yang berkelanjutan dituntut untuk lebih kreatif dan berinovasi. Kompetensi mahasiswa perlu dipersiapkan seiring dengan kebutuhan zaman saat ini. Dalam melawan perubahan budaya kerja, kemajuan teknologi dan sosial, mahasiswa memerlukan sistem yang mendukung sehingga dapat mengikuti perubahan di dunia industri yang kian cepat. Menciptakan kualitas sumber daya manusia, perguruan tinggi perlu mempersiapkan, mengimplementasikan kurikulum baru dan beragam.

Meningkatkan kemampuan agar sesuai bakat dan minat dilakukan secara komprehensif salah satunya dengan melakukan magang di perusahaan sebagai dasar pelatihan kerja. Harapan didapat generasi terbaik Indonesia bukan hanya impian. Magang di berbagai perusahaan diberbagai bidang, pertukaran mahasiswa dalam negeri, program wirausaha merdeka, dan lain sebagainya menjadi program Kampus Merdeka. Hal ini sesuai dengan kebijakan Kemendikbudristek Republik Indonesia (Kementerian Pendidikan Kebudayaan Riset dan Teknologi Republik Indonesia).

Untuk mempersiapkan keturunan terbaik anak Indonesia diperlukan program persiapan karir pemuda secara menyeluruh

yang merupakan unsur dari kebijakan Merdeka Belajar oleh Kemendikbudristek Republik Indonesia. Untuk mempersiapkan karir dimasa depan dengan memberikan kesempatan kepada mahasiswa dan mahasiswi untuk meningkatkan keahlian sesuai talenta dan kegemarannya dengan cara ikut serta ke dalam lingkungan kerja. Oleh karena itu harapannya, dalam setiap proses pembelajaran, suatu perguruan tinggi mampu mengembangkan inovasi-inovasi, yakni pembelajaran yang berpusat pada mahasiswa agar tercapainya lulusan yang berkualitas dan siap menghadapi perubahan zaman (Siregar et al., 2020).

Selasa (1/6/2021), dalam akun instagram Direktorat Jenderal Pendidikan Tinggi (Ditjen Dikti) Kemendikbud Ristek, Nadiem menjelaskan bahwa "Penerapan Kampus Merdeka untuk memastikan semua lulusan dapat berkarir sesuai dengan minat dan dengan adanya kampus merdeka mahasiswa dibekali disiplin ilmu guna menghadapi tantangan industri kerja". Program Kampus Merdeka memberikan kewenangan kepada mahasiswa untuk belajar tiga semester di luar kurikulumnya, saat 3 semester itu, mahasiswa selama 2 semester melaksanakan pembelajaran luar kampus yang mana nanti akan setara dengan 40 SKS didalam kelas dengan mengikuti silabus yang sama pada Perguruan Tinggi yang berbeda.

Dengan membuka peluang kepada mahasiswa untuk mengikuti Kampus Merdeka, maka akan memperluas dan meningkatkan wawasan dan kompetensi mereka di lingkungan kerja sesuai dengan minat, bakat dan cita-citanya.

Menko PMK (Menteri Koordinator Pembangunan Manusia dan Kebudayaan) Muhadjir Effendy memberikan penghargaan kepada Mendikbud atas buah pikiran “Merdeka Belajar“, dengan strategi ini pendidik menjadi lebih fokus di setiap pembelajaran siswa dan siswa pun bisa lebih banyak belajar” saat siaran pers kemdikbud 2019. Kita harus bersikap terbuka dan yakin saat menghadapi segala macam transformasi ini”.

Proses pembelajaran berkualitas bermakna harus dapat digunakan dalam menghadapi era disruptif. Dalam meraih prestasi mahasiswa dan dosen dengan kampus merdeka memberikan kebebasan sistem Pendidikan (Prahani et al., 2020) Kurikulum di Indonesia saat ini dianggap terlalu bersifat teoritis dan kurangnya kegiatan praktek praktek. Secara konseptual kebijakan MBKM sangat relevan dengan dunia kerja menjadikan mahasiswa sebagai pusat belajar di era revolusi industri 4.0. (Muslikh, 2020)

Dengan adanya kebijakan MBKM ini kampus melakukan revitalisasi kurikulum, sehingga tercipta pembelajaran yang sesuai otonom dan terwujudnya kultur belajar yang inovatif bebas, dan sesuai dengan kebutuhan siswa. Terciptanya pembelajaran sesuai dengan kurikulum ini memiliki tantangan (Hudjimartu et al., 2021) diantaranya:

1. Pembelajaran menggunakan system LMS (Learning Manajemen Sistem) untuk menunjang kegiatan luring maupun daring menggunakan server perguruan tinggi untuk segala bentuk kegiatan yang mengakibatkan permasalahan baru.
2. MoU kerjasama dengan perusahaan memungkinkan terjadi hambatan karena dikhawatirkan akan mengganggu proses produksi. (Muslikh, 2020)

Kampus merdeka memiliki program unggulan dimana salah satunya magang bersertifikat, dimana mahasiswa berpeluang besar diterima sebagai karyawan di tempat ia magang selain mendapat pengetahuan nyata dunia kerja. Saat ini program ini memberikan 2780 posisi untuk ditempati. Program Wirausaha Merdeka diselenggarakan 17 Perguruan Tinggi Negeri dan swasta dengan kuota 750 setiap kampus dimana memberi pendampingan *Business coach*. Program ini memberikan kesempatan mahasiswa mendalami studi yang diambil, sebagai wadah persiapan terjun di masyarakat. Berdasarkan data dari dari website resmi kemdikbud:

1. 725.000 lebih mahasiswa sudah terdaftar dan memiliki akun Kampus Merdeka.
2. 14.000 lebih Sekolah dari seluruh Indonesia telah menerima bantuan pengajaran
3. 3.000 lebih organisasi dan perusahaan telah mendaftarkan diri sebagai mitra
4. 1.300 lebih perguruan tinggi telah berpartisipasi diberbagai program Kampus Merdeka

Hal ini menunjukkan minat dari *stackholder* pada program yang sedang berjalan.

Analisis Sentiment termasuk dalam *text analytics* dimana tujuannya untuk memperoleh opini dari pengguna yang terdapat pada suatu *platform*. Data yang digunakan didapat dari berbagai sumber dari internet, media sosial salah satunya. Sosial media biasa digunakan untuk berburu mengumpulkan teman, bertukar informasi dan data dengan memanfaatkan web (Pasek et al., 2022) Teknik yang digunakan adalah *web scraping*. *Web scraping* yaitu suatu cara yang dapat dipergunakan untuk menghasilkan atau menyalin suatu big data dari dunia maya lalu dilakukan penyimpanan ke dalam file lokal komputer atau kedalam format tabel (*spreadsheet*) ((Rizaldi et al., 2017a). Sehingga dapat dilakukan pengolahan untuk mendapatkan informasi.

Analisis Sentimen menjadi salah satu bidang dari *Natural Language Processing*

atau biasa disingkat dengan NLP. NLP dipakai untuk membangun suatu sistem dengan mengidentifikasi dan menyalin opini dalam bentuk teks. NLP dan teks mining merupakan elemen dari *Artificial Intelligence* atau biasa dikenal dengan AI memungkinkan mengubah isi konten yang berbentuk teks menjadi sebuah data kuantitatif secara tepat, sehingga dapat digunakan sebagai pengambilan keputusan atau kebijakan. Analisis sentiment pada media sosial terdiri dari kumpulan data berupa tulisan atau text yang merepresentasikan banyak hal seperti emosi, opini, bahkan diskusi mengenai suatu topik hasil analisis sehingga nantinya dapat memberikan *insight* bagi organisasi.

Di era globalisasi informasi cepat berkembang sampai pada masyarakat umum dengan memanfaatkan berbagai macam aplikasi. Twitter sebagai salah satu jejaring media sosial yang populer di Indonesia. Menurut Statista per Januari 2022 pengguna aplikasi ini sebanyak 18,45 juta dan dilihat berdasarkan peringkat dunia, Indonesia berada pada peringkat ke 5. Opini melalui media sosial salah satunya berasal dari Twitter. Twitter dipandang kaya akan informasi sehingga dapat dijadikan sebagai sumber data dalam mengambil keputusan atau kebijakan. Twitter banyak dimanfaatkan oleh politikus, *intreprenur* dalam negeri dan mancanegara untuk mempromosikan produk, kebijakan politik bahkan berinteraksi dengan warga dan pejabat lainnya. Sebagian besar pemimpin dunia dan pejabat saat ini telah memiliki akun Twitter resmi. Analisa sentimen menghasilkan parameter yang bermanfaat untuk berbagai tujuan yang dapat jumpai pada ulasan, *feedback* ataupun sanggahan (Andreyestha & Azizah, 2022).

Twitter merupakan sumber informasi yang dapat digunakan dalam proses pengambilan keputusan dan analisa pandangan (sentimen) (Andreyestha & Azizah, 2020)). Twitter digunakan tidak hanya mengevaluasi strategi-strategi pemerintah namun melalui twitter dapat mencurahkan

keluhkesah atau mencurahkan pengalaman pada suatu produk atau jasa. Salah satunya pada strategi Kampus Merdeka. Sejak diluncurkannya kampus merdeka terdapat lebih dari 500 tweet dengan keyword Kampus Merdeka, berisi review maupun pertanyaan mengenai pendaftaran maupun pengalaman belajar.

Naive Bayes sebagai salah satu algoritma banyak dipakai dalam metode klasifikasi memiliki kinerja yang baik dan mudah dalam penerapannya (Fitriyani. Wahono R. S, 2015). Naive Bayes bukanlah metode yang baru digunakan untuk memprediksi segala kemungkinan yang terjadi pada suatu kasus. Naive Bayes merupakan salah metode klasifikasi data mining yang populer ((Fahdia et al., 2020). Algoritma ini memanfaatkan teori probabilitas, naive bayes merupakan bagian dari *supervised learning*, dimana menjadi salah satu algoritma yang paling cepat dalam mengolah beberapa fitur atau kelas. Pengklasifikasian pada naive bayes, memiliki ketidak bergantungan (*independence*) antara nilai atribut pada kelas dengan nilai atribut lainnya, disebut dengan *class-conditional independence* oleh karena itu proses pengolahannya dapat dilakukan dengan cara lebih sederhana sehingga disebut “naif (naive) (Han, 2012).

UNION Merupakan operator yang digunakan untuk menggabungkan input example sets sedemikian rupa sehingga semua atribut adalah bagian dari keseluruhan example sets.

Pada data mining dapat digunakan data latih yang memiliki atribut sentimen dan data uji yang tidak terdapat atribut sentimen. Dengan menggunakan operator union data latih dan data uji digabungkan kemudian data menghilangkan data yang berbeda. Dengan menggunakan metode ini memiliki keunggulan pada analisa yang memiliki banyak data.

Synthetic Minority Oversampling Technique (SMOTE) merupakan metode *oversampling* dimana berfungsi untuk

memecahkan permasalahan ketidakseimbangan pada kelas. Pertama dilakukan pendekatan mengklasifikasi dari sekelompok data yang tidak seimbang. Dimana kategori klasifikasi tidak disajikan secara seimbang. Kedua, algoritma yang terdapat pada SMOTE akan mengurangi nilai selisih antara vektor dari fitur pada kelas minoritas dan nilai nearest neighbor dari kelas minoritas lalu dikalikan dengan nilai acak antara 0 sampai 1. Kemudian, menambahkan hasil perhitungan dengan vektor fiturnya maka didapatkan hasil nilai vektor yang baru (Jishan et al., 2015).

$$(X_{\text{new}} = X_i + (X_i^{\wedge} - X_i) \times \delta) \quad (2)$$

Keterangan :

X_i = vektor dari fitur pada kelas minoritas

X_i^{\wedge} = k-nearest neighbors untuk X_i

δ = angka acak antara 0 sampai 1

Algoritma Naïve Bayes digunakan dalam mendukung penelitian ini dengan menggunakan metode klasifikasi. Pada analisa, dilakukan dengan mencocokkan beberapa kombinasi algoritma yang akan diuji pada twitter dengan menggunakan kata kunci Kampus Merdeka. Besarnya data merupakan salah satu tantangan pada analisa sentiment, sehingga memerlukan algoritma Union untuk mengoptimalkan hasil dan *Synthetic Minority over Sampling Technique* sebagai penyeimbang dan mengetahui keakuratan data. Data dikelola dengan menggunakan aplikasi Rapidminer.

1.1 Penelitian Terkait

Banyak penelitian tentang analisis sentimen dengan menggunakan metode *crawling* dengan pengumpulan data pada media sosial maupun aplikasi yang terdapat dimana dapat digunakan sebagai review atau ulasan dan pengambilan keputusan bagi masyarakat atau produk yang diulas.

Penelitian dilakukan oleh (A. Andreyestha & Azizah, 2022) dimana meneliti kicauan twitter pada salah satu marketplace Tokopedia. Bukan hanya ulasan produk, komplain pelanggan, tetapi juga pengalaman dan kepuasan dalam menggunakan marketplace tersebut. Penelitian ini menggunakan algoritma Naïve Bayes dan Random Forest untuk memberikan hasil klasifikasi yang diharapkan, analisis akan dilakukan dengan membandingkan beberapa kombinasi algoritma yang akan diuji pada kicauan twitter mengenai tokopedia, dengan menggunakan kombinasi Synthetic Minority Oversampling Technique (SMOTE) untuk mengoptimalkan data yang tidak seimbang. Sehingga menghasilkan pengujian kicauan twitter tokopedia SMOTE dapat meningkatkan akurasi algoritma Naive Bayes menjadi 86,93% meningkat 3,4%.

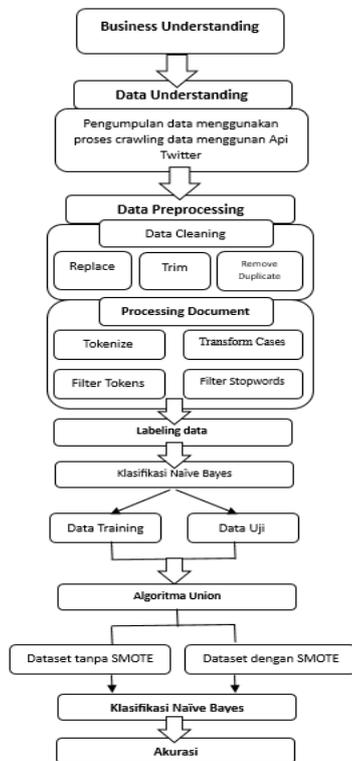
Pada penelitian (Fitriyani. Wahono R. S, 2015). membahas tentang pengujian cacat software dengan menggunakan metode Naive Bayes untuk prediksi cacat software yaitu menentukan software mana yang masuk kategori cacat dan tidak cacat. Data yang digunakan merupakan dataset publik yang berasal dari NASA MDP. Untuk menangani ketidakseimbangan kelas dalam dataset NASA MDP menggunakan metode ensemble (bagging), bagging salah satu bagian dari metode ensemble untuk memperbaiki ketidakseimbangan kelas. Sedangkan untuk menangani data yang berdimensi tinggi atau fitur-fitur yang tidak memiliki kontribusi dengan menggunakan seleksi fitur greedy forward selection. Hasil dalam penelitian ini didapatkan t dari rata rata peringkat bahwa naïve bayes dengan greedy forward selection dan bagging merupakan model yang terbaik dalam prediksi cacat software.

(Ni'ma Kholila, 2021) menganalisa sentimen program merdeka belajar pada Kampus Merdeka dengan menggunakan data Twitter. Metode klasifikasi Support Vector Machine

diterapkan untuk memprediksi sentimen positif dan negatif dari cuitan di twitter. Menggunakan Python Library digunakan untuk menghitung nilai precision, recall, dan F-measure untuk masing-masing kelas (kelas positif maupun kelas negatif). Precision bernilai tinggi menunjukkan kecilnya nilai False Positive (ter prediksi positive, sebenarnya negative). Nilai recall yang tinggi menunjukkan kecilnya nilai False Negative (ter prediksi negative, sebenarnya positive). Nilai F-measure yang baik akan mengindikasikan bahwa model klasifikasi sentimen mempunyai precision dan recall yang baik. Sehingga dapat disimpulkan bahwa nilai akurasi untuk masing-masing kelas dilihat dari nilai F-measure.

2. Metode Penelitian

Guna mendukung penulisan maka diperlukan kerangka penelitian yang merupakan tahapan dari proses penelitian. Tahapan penelitian yang sesuai dengan penelitian sebagai berikut:

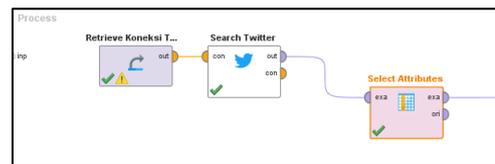


Gambar 1. Tahapan Penelitian

Uraian kerangka penelitian sebagai berikut:

a. Data *understanding*

Pengumpulan data menggunakan proses *crawling* data menggunakan api Twitter dimana data berupa data text, tweet yang diambil menggunakan keyword “kampus merdeka”. *Scraping* data dilakukan dengan menarik tweet pada bulan Januari 2022 sampai dengan Agustus 2022. Dimana jumlah data ada 680, sedangkan data yang tidak diberi label berjumlah 430 yang merupakan data raw yang berisi ungkapan, pertanyaan, pengalaman. Kemudian data disimpan di file lokal sehingga dapat dilakukan proses analisa dengan menggunakan aplikasi Rapidminer yaitu *preprocessing*.



Gambar 2. Proses Pengumpulan Data

Setelah pengumpulan data dilakukan menggunakan *Search twitter* kemudian menggunakan *select atribut* untuk menampilkan atribut yang diinginkan atau yang diperlukan saja. Berikut contoh data tweet:

Row No.	Text
1	@collegemertless Ga semua malfut tersedia emang nnder. Tesserah kebijakan si kampus malfut itu bisa dipake buat program merd...
2	umil Disaini ada yang ngalamin nison di alun kampus merdeka udah terverifikasi tapi nison nya Gaada di Diiri. Aku harus apa ya soalnya...
3	@FIBES3 Coba ihut kampus merdeka kait
4	Kegiatan Pembekalan Program Alumni Kampus Merdeka day-1
5	Halo teman2👋🏻
6	@Idafalala ihut kampus merdeka yg lainnya aja leh
7	Nadlem menekankan, lewat anggaran tersebut salah satu prioritas Mendikbud adalah menambah program pusat talenta. #Rumpara...
8	Nadlem menekankan, lewat anggaran tersebut salah satu prioritas Mendikbud adalah menambah program pusat talenta. #Rumpara...
9	@Idafalala Tapi kalo bole beropini hehe. ihutnya yang beda program leh. kan kampus merdeka menunya macem macem tauu
10	@Idafalala Soalnya stripisi itu komitmen aja. kampus merdeka usia...

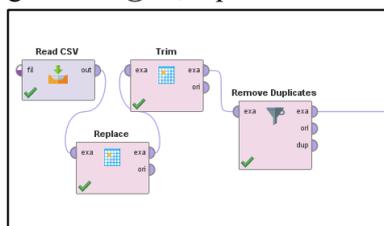
Gambar 3. Contoh Data Tweet

b. Data *Preprocessing*

Pada tahapan *preprocessing* mencakup beberapa kegiatan pembersihkan data hingga diharapkan memiliki hasil data yang akurat kemudian diproses ke langkah

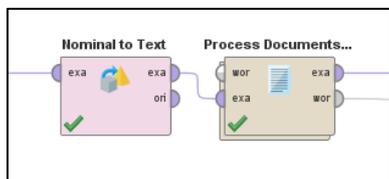
berikutnya. Tahapan pra proses data meliputi meliputi *cleaning*.

1. *Replace*, dapat digunakan berkali-kali secara manual agar mendapatkan hasil yang baik. Proses *replace* berguna untuk mengekstrak informasi dengan menghapus kata-kata dan simbol yang tidak berguna dengan menggunakan *regular expression* atau biasa disebut *regress RT @*. ,http.**.



Gambar 4. Proses Pembersihan

2. *Trim* digunakan untuk menghilangkan tanda spasi yang tidak diperlukan.
3. *Operator Remove Duplicates* digunakan sebagai penghilangan duplikasi data misal hasil dari retweet.



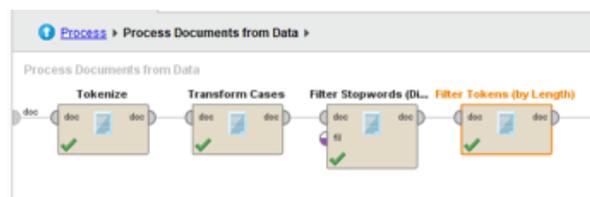
Gambar 5. Pemrosesan Dokumen

Langkah awal pemrosesan data dengan mengubah atribut nominal menjadi atribut text. Setiap nilai nominalnya hanya digunakan sebagai nilai string dari atribut baru. *Process documents* memiliki beberapa subproses di dalamnya, yaitu:

- b. *Tokenize*

Suatu metode untuk membagi teks menjadi bagian yang lebih kecil (kata, kalimat, bigrams) (Rizaldi et al., 2017b). Memisahkan satu kalimat menjadi beberapa kata sehingga kata satu kata memiliki atribut sendiri. Menghasilkan data TF-IDF yaitu teknik untuk menghitung suatu kata muncul dalam satu dokumen.

- c. *Transform Cases* digunakan untuk menyamakan tipe teks menjadi huruf kecil semua atau besar semua.
- d. *Filter Stopwords Dictionary*, berfungsi menghapus kata-kata yang tidak penting, karena data yang diambil menggunakan Bahasa Indonesia maka menggunakan *dictionary*. Dimana list kata diambil dari *Kaggle.com* berbahasa Indonesia
- e. *Filter Tokens by Length* untuk memfilter kata-kata yang jumlah karakternya terlalu sedikit atau terlalu banyak dimana secara default minimal karakter 4 dan maksimal karakter 25.



Gambar 6. Preprocessing Data

Setelah membentuk model seperti diatas pada *Rapidminer*, maka akan muncul kalimat yang baku dan lebih sederhana.

- c. *Klasifikasi Naïve Bayes dan Algoritma Union*.

Setelah preprocessing selesai mulailah pengolahan data. Dimana sumber data awal berjumlah 655 data setelah dilakukan pembersihan menggunakan operator *Remove Duplicate* membuang kata-kata yang berulang dan menghapus kata-kata yang tidak mengandung sentiment maka dilakukan pelabelan secara manual. Berikut data yang diproses selanjutnya. Setelah data dibersihkan menghasilkan:

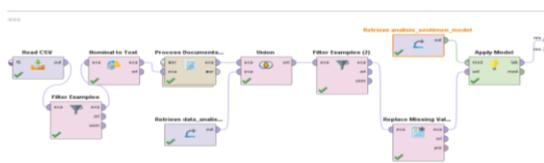
The screenshot shows the 'ExampleSet (Remove Duplicates)' window in Orange software. It displays a table with columns for Name, Type, Missing, and Statistics. The table has two rows: 'Text' and 'sentimen'.

Name	Type	Missing	Statistics	Filter (2/2 attributes)
Text	Polynomial	0	Least yang per [] juga (1) Most Alhamdul [] alon (5)	Value Alhamdul [] an kallen (5), Alhamdul [] n kalle
sentimen	Polynomial	195	Least positif (1) Most positif (372)	Value positif (372), negatif (56), [1 more]

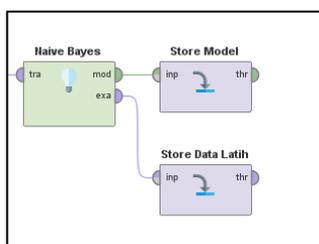
Gambar 9. Pembersihan Data

Berdasarkan gambar 9. menghasilkan data positif sebanyak 372 dan negatif 58 sisanya yaitu 195 data belum diberi label. Karena data yang banyak maka tidak mungkin melakukan memberikan labelan secara manual secara keseluruhan. Oleh karena itu diperlukan proses selanjutnya yaitu pengolahan data dengan operator Union dan Naïve Bayes.

Pada proses ini dengan menggunakan algoritma Naïve Bayes yang digunakan sebagai mechine learning. Dengan algoritma ini akan mempelajari pola data dari kata-kata yang ada sehingga akan menghasilkan data model klasifikasi sehingga dapat dipakai Analisa sentiment dari tweet yang belum memiliki nilai sentiment. Pada proses ini model dan data disimpan untuk proses selanjutnya.



Gambar 10. Proses Algoritma Union



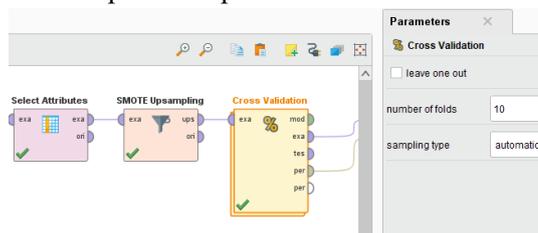
Gambar 11. Model dan Data Latih Klasifikasi

Berdasarkan model diatas, akan didapatkan prediksi positif atau negative dari algoritma Naïve Bayes dan penggabungan data Union

d. Pengolahan Data SMOTE

Perbandingan antara sentiment positif dan negative yang terlalu jauh akan menimbulkan *imbalance class*. Fenomena *imbalance class* akan menyebabkan penurunan kinerja klasifikasi. Untuk mengatasi permasalahan *imbalance class*,

salah satu metode yang digunakan adalah sampling (Sutoyo et al., 2020). Salah satu teknik over-sampling yang dapat digunakan adalah *Syntetic Minority Over-sampling Technique* (SMOTE). Dengan menggunakan menggunakan *Syntetic Minority Over Sampling Technique* (SMOTE) penulis gunakan untuk mengetahui optimasi dan keakuratan data. Data sentiment yang sebelumnya masih kosong digantikan dengan data yang sudah diproses di proses union.



Gambar 14. Proses SMOTE dan Validasi

Operator validasi tersebut memiliki subproses untuk mendapatkan hasil nilai evaluasi. Sub proses pada operator *cross validation* dapat dilihat pada gambar berikut:



Gambar 15. Subproses dalam Cross Validation

3. Hasil dan Pembahasan

3.1. Hasil *Preprocessing*

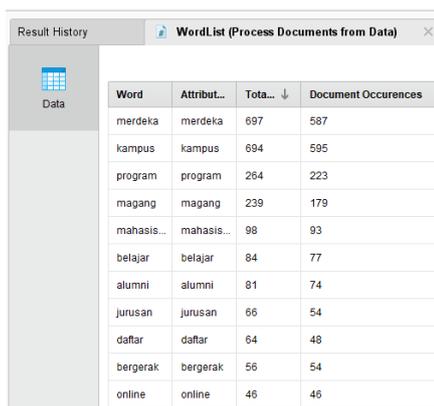
Berikut merupakan contoh hasil pemrosesan dokumen

Tabel 1 Hasil pemrosesan dokumen

Data Awal	Data hasil pemrosesan
Yh ak ikut magang kampus merdeka dpt uang saku kerja berangkat suka-berangkat suka-suka tapinya tau diri si	Magang kampus merdeka uang saku kerja berangkat suka suka tapinya

Kalau anak pns gak bisa dapat keringanan ukt setauku tapi misal beasiswa bisaa bener program kampus merdeka dapat uang saku bantuan ukt Semangat nderr????	Anak dapat keringanan setauku beasiswa bisaa bener program kampus merdeka dapat uang saku bantuan semangat
Kangen ga kalian sama masa masa belajar itu Btw skrg KMMI ko gaada lagi ya hmm kasian adek tingkat mau nyoba malah gaada	Kangen belajar skrg kmmi gaada kasian adek tingkat nyoba gaada
Dokumen yang diperluin buat kampus merdeka apa aja sih soalnya blm input dokumen dan lagi nungguin surat rekomendasinya jadi makanya blm bisa liat halaman	Dokumen diperluin kampus merdeka input dokumen nungguin surat rekomendasinya liat halaman

Untuk mengetahui data apa yang sering muncul dapat diketahui dengan proses *WordList to data*. Sehingga menghasilkan kata-kata penting yaitu merdeka, program, kampus dan magang dan mahasiswa.



Word	Attribut...	Tota... ↓	Document Occurrences
merdeka	merdeka	697	587
kampus	kampus	694	595
program	program	264	223
magang	magang	239	179
mahasis...	mahasis...	99	93
belajar	belajar	84	77
alumni	alumni	81	74
jurusan	jurusan	66	54
daftar	daftar	64	48
bergerak	bergerak	56	54
online	online	46	46

Gambar 7. List Kata Penting

Dengan menggunakan operator *Worldcloud* menampilkan atau memvisualisasikan kata yang sering muncul, seperti dibawah ini:



Gambar 8. Hasil dari Worldcloud

Berdasarkan gambar diatas, kata yang sering muncul pada data twitter yaitu program kampus merdeka, setelah itu kampus merdeka lalu magang kampus merdeka.

3.2. Hasil Pengolahan Naïve Bayes dan UNION

Filter Example berguna untuk mengambil data yang belum memiliki nilai sentiment. Data yang belum memiliki sentiment disebut data uji akan digabungkan atributnya dengan data latih dengan menggunakan algoritma Union sehingga mempunyai irisan data. Setelah data digabung dengan menggunakan operator union maka data sample akan bertambah begitu juga atributnya. Filter Example membuang data latih yang berbeda hal ini menyebabkan beberapa kolom kosong atau disebut *missing value*. Proses ini menghasilkan:

Tabel 2. Hasil Operasi Union

Row No.	Senti men	Prediction	text	confirm
12	?	negatif	produk kampus merdeka merdeka memaki kasar kotor memfitnah	0
13	?	negatif	produk kampus merdeka merdeka memaki berkata kasar memfitnah	0

14	?	positif	akibat kampus merdeka
15	?	positif	kampus merdeka membuatkumorning person jompo mahasiswa
16	?	positif	petukaran pelajar kampus merdeka

Dari proses pengolahan menggunakan metode SMOTE, menghasilkan informasi keakurasian data sebagai berikut:

Tabel 4. Perhitungan akurasi menggunakan SMOTE
 accuracy: 91.21% +/-2.71% (micro average:91.21%)

	true positif	true negatif	class precision
pred.positif	403	0	100.00%
pred.negatif	86	489	85.04%
class recall	82.41%	100.00%	

Pada gambar terlihat kolom sentiment yang kosong, kolom prediksi yang merupakan hasil dari proses prediksi algoritma Naïve Bayes dan penggabungan data Union dan kolom text yang berisi kalimat yang diprediksi.

3.3. Hasil Pengolahan Metode SMOTE

Tingkat akurasi menggunakan *confussion matrix* setelah dilakukan pengolahan menggunakan naïve bayes dan union tanpa menggunakan metode SMOTE, yaitu sebesar 77,60%.

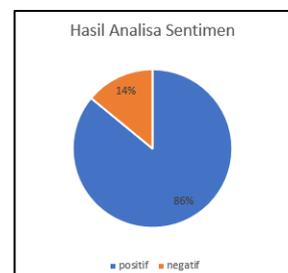
Tabel 3. Hasil Akurasi Data *confussion matrix*

accuracy: 77.58% +/-7.95% (micro average:77.60%)			
	true positif	true negatif	class precision
pred.positif	405	43	90.40%
pred.negatif	84	35	29.41%
class recall	82.82%	44.87%	

Berdasarkan gambar diatas, terdapat 489 data, dimana 405 data memberikan hasil prediksi *positif* yang sesuai (benar) dengan *positif*, kemungkinan 43 *positif* ternyata *negatif*, kemungkinan 84 data *negatif* ternyata *positif*, dan kemungkinan 35 data *negatif* sesuai dengan *negatif*. Dengan persentase kemungkinan yang positif sebesar 90,40% dan persentase kemungkinan yang negatif sebesar 29,41%. Persentase *class recall* pada yang positif sebesar 82,82% dan persentase *class recall* yang benar negatif sebesar 44,87%.

Berdasarkan gambar diatas, diketahui dari 489 data didapat kemungkinan dari 403 data *positif* sesuai real dengan *positif*, 0 *positif* ternyata *negatif*, kemungkinan 86 data *negatif* ternyata *positif*, dan kemungkinan 489 data *negatif* sesuai dengan *negatif*. Dengan persentase kemungkinan yang positif sebesar 100% dan persentase kemungkinan yang negatif sebesar 85,04%. Persentase *class recall* pada yang positif sebesar 82,41% dan persentase *class recall* yang benar negatif sebesar 100%.

Hasil dari proses analisa sentiment Kampus Merdeka yang telah dilakukan menghasilkan keakuratan menggunakan operator SMOTE 91,21%. Sehingga selisih tanpa menggunakan SMOTE cukup besar yaitu 13.61%.



Gambar 16. Presentasi Hasil Analisa Sentimen

Besarnya minat yang besar pada Kampus Merdeka dapat dilihat dari akumulasi sentiment yang bernilai positif dan negative

menghasilkan 86% bernilai positif dan 14% bernilai negative. Berdasarkan sumber data 680 sentimen. Hal ini sesuai dengan analisis static deskriptif yang dilakukan oleh (Suyoto Ario, Ugung Dwi et al, n.d.) dimana peningkatan *softskill* mahasiswa peserta Kampus Merdeka sangat tinggi yaitu 4,48 dari skala likert lima tingkat. Rata-rata minat mahasiswa mengikuti Kampus Merdeka juga sangat tinggi yaitu 4,37 dengan skala likert lima tingkat.

Program MBKM memberikan banyak keuntungan bagi mahasiswa agar dapat mendalami ilmu yang diminati sebagai persiapan terjun di dunia kerja. Program ini sangat baik namun tidak luput dari kekurangan seperti Sistem informasi akademik belum siap, SDM sebagai eksekutor program ini belum cukup kuat. Namun kebijakan Kampus Merdeka sangat bermanfaat bagi mahasiswa dan dosen sehingga partisipasi akan meningkat demikian juga tingkat kepuasan ungkap Rektor Ukrida Wani Devita Gunardi, 2021.

4. Simpulan

Program Kampus Merdeka memberikan pengalaman positif bagi mahasiswa untuk belajar dengan memberikan keluasaan ilmu pengetahuan dan pengalaman pada dunia kerja sehingga banyak diminati mahasiswa. Berdasarkan hasil pengkajian sebelumnya, dapat disimpulkan bahwa analisis sentimen menggunakan algoritma Union memudahkan proses pengolahan data yang berjumlah besar. Menggunakan algoritma Naïve Bayes dapat memprediksi suatu sentiment bernilai positif atau negative. Berdasarkan hasil analisa akumulasi sentiment menghasilkan nilai positif lebih besar yaitu 86% dibandingkan yang memberikan nilai negative yaitu 14%. Berdasarkan data Program Kampus Merdeka memiliki opini yang baik. Tingkat keakurasian data tanpa menggunakan algoritma SMOTE yaitu 77.16%, dengan

mengintegrasikan SMOTE dan algoritma Naïve Bayes dari kicauan dengan kata kunci Kampus Merdeka menghasilkan tingkat keakurasian meningkat sebesar 91.21%. Sehingga teknik Smote dapat meningkatkan keakurasian.

Kebijakan publik lainnya mungkin dapat dilakukan Analisa sentiment sehingga dimanfaatkan sebagai bahan penilaian maupun evaluasi. Penggunaan teknik klasifikasi lain masih sangat mungkin untuk digunakan. Guna memperoleh nilai akurasi yang lebih baik, penelitian selanjutnya dapat menggunakan lebih banyak data set. Sumber data dapat diambil dari media social yang lain.

5. Referensi

- Andreyestha, A., & Azizah, Q. N. (2022). Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma Smote. *Infotek : Jurnal Informatika Dan Teknologi*, 5(1), 108–116.
<https://doi.org/10.29408/Jit.V5i1.4581>
- Andreyestha, Q. N. A. (N.D.). *Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan algoritma Smote*.
- Fahdia, M. R., Asnonik, N., & Putri, S. A. P. (2020). Implementasi Metode Naive Bayes Pada Analisa Penerimaan Kartu Jakarta Pintar Di Smk Al Kautsar Jakarta. *Inti Nusa Mandiri*, 14(2), 231–236.
<https://doi.org/10.33480/Inti.V14i2.1779>
- Fitriyani, Wahono R. S. (2015). Romi Id-Integrasi-Bagging-Dan-Greedy-Forward. *Integrasi Bagging Dan Greedy Forward Selection Pada Prediksi Cacat Software Dengan Menggunakan Naive Bayes*, 1(2).
- Han, J. Dan M. Kamber. 2012. *Data Mining Concepts And Techniques Third*

- Edition. San Francisco: Morgan Kaufmann.
- Hudjimartu, S. A., Prayudyanto, M. N., Permana, S., & Heryansyah, A. (2021). Peluang Dan Tantangan Implementasi Merdeka Belajar Kampus Merdeka (Mbkm) Di Fakultas Teknik Dan Sains Uika Bogor. *Educate : Jurnal Teknologi Pendidikan*, 7(1), 58. <https://doi.org/10.32832/educate.v7i1.6245>
- Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving Accuracy Of Students' Final Grade Prediction Model Using Optimal Equal Width Binning And Synthetic Minority Over-Sampling Technique. *Decision Analytics*, 2(1). <https://doi.org/10.1186/s40165-014-0010-2>
- Muslikh. (2020). Landasan-Filosofis-Dan-Analisis-Terhadap Kebijakan Mbkm. *Jurnal Syntax Transformation*, 1(3), 40–46.
- Ni'ma Kholila. (2021). Merdeka Belajar Twitter. *Jurnal Ilmiah Teknik Informatika*, 15(2), 252–261. <https://doi.org/10.35457/antivirus.v15i2.1866>
- Pasek, P., Mahawardana, O., Sasmita, G. A., Agus, P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini Dari Media Sosial Twitter Terhadap “Figure Pemimpin” Menggunakan Python. In *Jitter-Jurnal Ilmiah Teknologi Dan Komputer* (Vol. 3, Issue 1).
- Pustakalewi.Com (2022, 31 December). Rektor Ukrida: Kebijakan Merdeka Belajar Kampus Merdeka Bermanfaat Bagi Dosen Dan Mahasiswa. Rektor Ukrida: Kebijakan Merdeka Belajar Kampus Merdeka Bermanfaat Bagi Dosen Dan Mahasiswa - Pustakalewi News
- Prahani, B. K., Deta, U. A., Yasir, M., Astutik, S., Pandiangan, P., Mahtari, S., & Mubarak, H. (2020). The Concept Of “Kampus Merdeka” In Accordance With Freire’s Critical Pedagogy. *Studies In Philosophy Of Science And Education (Sipose)*, 1(1), 21–37. <http://sciejournal.com/index.php/sipose>
- Rizaldi, T., Teknologi, J., Politeknik, I., Jember, N., & Arief, H. (2017a). Perbandingan Metode Web Scraping Menggunakan Css Selector Dan Xpath Selector. 43 *Teknika*, 6(1). <http://blog.detik.com/>
- Rizaldi, T., Teknologi, J., Politeknik, I., Jember, N., & Arief, H. (2017b). Perbandingan Metode Web Scraping Menggunakan Css Selector Dan Xpath Selector. 43 *Teknika*, 6(1). <https://doi.org/10.34148/teknika.v6i1.56>
- Siregar, N., Sahirah, R., & Harahap, A. A. (2020). Konsep Kampus Merdeka Belajar Di Era Revolusi. *Fitrah: Journal Of Islamic Education*, 1(1), 141–157. <http://jurnal.staisumatera-medan.ac.id/index.php/fitrah>
- Sutoyo, E., Asri Fadlurrahman, M., Telekomunikasi Jl Terusan Buah Batu, J., Dayeuhkolot, K., Bandung, K., & Barat, J. (2020). Jepin (Jurnal Edukasi Dan Penelitian Informatika) Penerapan Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi Dan Penelitian Informatika*, 6(3), 379–385.
- Suyoto Ario.Ugung Dwi Et Al. (N.D.). *Analisis Persepsi Minat Mahasiswa Berpartisipasi Dalam Program Mbkm Dalam Meningkatkan Softskill Lulusan* (Vol. 21). Online.