

Penerapan Text Mining Menggunakan Algoritme Naïve Bayes Dalam Mengklasifikasi Sentimen Netizen di media sosial *Twitter* (Studi Kasus Pertemuan KTT G20 di Indonesia)

Yusup Yuliadi^{*1}, Magdalena A. Ineke Pakereng²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail *Corresponding Author*: 672019317@student.uksw.edu

Abstract

The G20 is a forum that discusses global issues including Finance Track and Sherpa Track, this forum consists of developed and developing countries. The purpose of this research is to classify positive and negative tweets using the naive bayes method. The naive bayes method can predict future possibilities based on past experience and this method is recommended from several previous researchers because this method is considered suitable for analyzing positive and negative tweets in the dataset. The existing data is processed using rapidminer and produces a model. Based on the model of the naive bayes method seen from 700 training data, 675 positive predictions with class precision 100.00% and 22 negative predictions with class precision 88.00% and with an analysis accuracy value of 99.57%. So it can be concluded that many people have positive sentiments on twitter when the G20 is held in Indonesia and the G20 can be an effort to encourage the country's economy to be even better.

Keyword: *Text mining; G20 Summit; Naive Bayes; Netizen sentiment*

Abstrak

G20 merupakan forum yang membahas isu masalah global antara lain, *Finance Track* dan *Sherpa Track*, forum ini beranggotakan negara maju dan berkembang, Tujuan penelitian ini adalah mengklasifikasikan *tweet* positif dan negatif menggunakan metode *naive bayes*. Metode *naive bayes* dapat memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu serta metode ini direkomendasikan dari beberapa peneliti sebelumnya karena metode ini dianggap cocok untuk menganalisa *tweet* positif dan negatif dalam *dataset*. Data yang ada diolah menggunakan *rapidminer* dan menghasilkan model. Berdasarkan model dari metode *naive bayes* dilihat dari 700 data latih, 675 prediksi positif dengan *class precision* 100.00% serta 22 prediksi negatif dengan *class precision* 88.00% dan dengan nilai analisa *accuracy* 99,57%. Jadi dapat disimpulkan banyak masyarakat yang bersentimen positif di twitter saat digelarnya G20 di Indonesia dan G20 ini dapat menjadi upaya untuk mendorong perekonomian negara agar lebih baik lagi.

Kata kunci: *Text Mining; Konferensi Tingkat Tinggi G20; Naive Bayes; Sentimen Netizen*

1. Pendahuluan

Latar belakang pembentukan G20, atau *Group of Twenty*, terjadi karena pecahnya krisis keuangan Asia pada tahun 1998, ketika deklarasi tersebut menjadi forum bagi negara maju dan negara berkembang untuk membahas isu-isu penting yang mempengaruhi ekonomi dunia. G20 mendorong kerja sama untuk mencapai pertumbuhan ekonomi global yang stabil, G20 merupakan sebuah forum yang melakukan kerja sama multilateral atau kerja sama yang dilakukan untuk mendapatkan kepentingan tertentu, yang terdiri dari 19 negara utama dan Uni Eropa (UE) yang memiliki pemasukan atau pendapatan menengah hingga tinggi, atau biasa disebut dengan negara maju dan berkembang. G20 biasanya membahas isu global yakni *Finance Track* yang berfokus untuk membahas isu keuangan antara lain; kebijakan fiskal, moneter dan rill, investasi infrastruktur, regulasi keuangan, inklusi keuangan, perpajakan internasional. Biasanya pertemuan untuk membahas *Finance Track* dihadiri oleh menteri

keuangan atau gubernur bank sentral dari masing-masing anggota. Kemudian ada *Sherpa Track* isu ini khusus dalam membahas diluar keuangan, antara lain; anti korupsi, ekonomi digital, lapangan kerja, pertanian, pendidikan, urusan luar negeri, budaya, kesehatan, pembangunan, lingkungan, pariwisata, energi berkelanjutan, perdagangan investasi dan industri, dan pemberdayaan perempuan. Untuk pertemuan yang membahas *Sherpa Track* ini biasanya dihadiri oleh para delegasi dari berbagai negara anggota [1]-[3].

Pertemuan KTT G20 pada November 2022 diselenggarakan di Indonesia. Di mana Indonesia terpilih sebagai tuan rumah pada pertemuan ini yang membahas berbagai agenda untuk pertumbuhan ekonomi, diantaranya *Business 20 (B20)*. B20 ini adalah forum yang membahas bisnis atau pasar internasional dan B20 ini dikhususkan untuk pebisnis swasta dalam menggerakkan pertumbuhan ekonomi yang lebih kuat lagi dimasa mendatang. Pertemuan KTT G20 menimbulkan kontroversial di kalangan masyarakat yang menyebabkan munculnya berbagai opini warganet khususnya masyarakat Indonesia yang dipaparkan melalui postingan, komentar dan tanggapan di media sosial salah satunya yakni twitter [2], [4], [5].

Perlunya dilakukan penilaian pelaksanaan kegiatan KTT G20 dengan melihat tanggapan (sentimen) masyarakat mengenai pelaksanaan kegiatan tersebut. Proses penilaian dan pemaparan hasil sentimen masyarakat dapat berjalan efisien dan efektif maka, diperlukan suatu model komputasi untuk memperoleh hasil yang jitu serta mempercepat proses pemaparan hasil jika, dilakukan secara manual.

Dalam bidang Data Mining, terdapat beberapa model algoritme komputasi yang dapat digunakan untuk melakukan analisis data, salah satunya adalah algoritme Naïve Bayes. Algoritme Naïve Bayes merupakan sebuah metode klasifikasi yang menggunakan teknik supervised untuk memprediksi label kelas dari instance dengan menggunakan probabilitas bersyarat. Algoritme ini telah diuji dan diterapkan dalam berbagai bidang, termasuk bidang kesehatan, kebijakan publik, dan berbagai bidang bisnis lainnya [6].

Penelitian ini dilakukan untuk mengklasifikasikan sentimen positif dan negatif di media sosial twitter tentang pertemuan KTT G20 di Indonesia. Hasil penelitian ini diharapkan bisa digunakan pemerintah untuk menjadi acuan pengambilan keputusan dalam menyelenggarakan kegiatan kenegaraan berikutnya dan membantu pemerintah untuk menjelaskan kepada masyarakat bahwa penyelenggaraan G20 di Indonesia bertujuan untuk menyejahterakan masyarakat.

2. Tinjauan Pustaka

Penelitian terdahulu yang berjudul "Analisa Sentimen Warganet Terhadap KTT G20 Indonesia Menggunakan Algoritme Naïve Bayes" penelitian ini membahas bagian dari evaluasi pelaksanaan KTT G20 di Indonesia, penting bagi pemerintah Indonesia untuk menerima masukan dari berbagai pihak, termasuk menganalisis sentimen masyarakat terhadap acara tersebut melalui penggunaan Twitter. Tujuan dari analisis ini adalah untuk mengklasifikasikan tanggapan masyarakat menjadi positif dan negatif menggunakan metode *Naïve Bayes Classifier (NBC)*. Proses analisis melibatkan pengumpulan data dari Twitter, pra-pemrosesan data, klasifikasi menggunakan Naïve Bayes, validasi hasil, dan evaluasi. Hasil pengujian menunjukkan bahwa kinerja *Naïve Bayes* mencapai akurasi sebesar 94,55%, presisi 100%, dan recall 89,10%. Informasi ini diharapkan dapat membantu pemerintah dalam pengambilan keputusan terkait kerjasama G20 di masa depan [6].

Penelitian yang dilakukan oleh Kevin, Bayu, dan Dian yang berjudul "Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritme *Backpropagation* dan *Term Frequency-Inverse Document Frequency*" yang bertujuan untuk menganalisis pendapat masyarakat terkait Undang-Undang Cipta Kerja (UU Cipta Kerja) menggunakan data dari Twitter. Peneliti menggunakan Jaringan Syaraf Tiruan (Artificial Neural Network) dengan algoritme *Backpropagation* dan metode pembobotan term TF-IDF untuk mengklasifikasikan tweet menjadi sentimen positif atau negatif. Hasil analisis *word cloud* juga mengungkapkan kesamaan dan perbedaan dalam kedua sentimen, dengan kata-kata seperti pemerintah, buruh, pekerja, rakyat, dan masyarakat menjadi yang paling sering muncul. Penelitian ini memberikan wawasan yang berharga tentang pandangan masyarakat terhadap UU Cipta Kerja berdasarkan data dari Twitter [7].

Penelitian yang serupa dengan judul "Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode *Fined-Grained Sentiment Analysis*" yang membahas Media

sosial, terutama Twitter, telah memiliki pengaruh yang signifikan dalam membentuk opini, pandangan, sentimen, dan preferensi politik publik selama Pemilihan Umum. Penelitian ini bertujuan untuk mengidentifikasi percakapan yang terjadi di Twitter selama debat pertama calon presiden Republik Indonesia dengan menggunakan hashtag dari kedua pasangan calon. Selain itu, penelitian ini juga bertujuan untuk memahami kecenderungan sentimen masyarakat di Twitter terkait dengan debat tersebut, apakah cenderung positif, negatif, atau netral. Data percakapan di *Twitter* diperoleh melalui Twitter API menggunakan bahasa pemrograman R. Metode analisis sentimen yang digunakan adalah *Fined-grained Sentiment Analysis*. Hasil penelitian menunjukkan bahwa tweet dengan menggunakan kedua hashtag cenderung memiliki sentimen positif yang lebih dominan daripada sentimen negatif dan netral [8].

Penelitian terdahulu yang berjudul “Klasifikasi dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritme *Naïve Bayes*. (Studi Kasus: Pekan Olahraga Nasional XX 2021)” yang membahas analisis sentimen terkait Pekan Olahraga Nasional (PON) XX 2021 menggunakan data dari Twitter, dengan jumlah data tweet sebanyak 218 yang belum terlabelkan. Tahapan penelitian meliputi pemberian label, pengolahan teks, dan pemberian bobot kata menggunakan metode Term *Frequency-Inverse Document Frequency* (TF-IDF). Untuk mengatasi ketidakseimbangan data, digunakan Synthetic Minority Oversampling Technique (SMOTE). Selanjutnya, data dibagi menjadi data latih dan data uji, dan diklasifikasikan menggunakan metode *Naïve Bayes*. Dari penelitian ini, dapat disimpulkan bahwa data tweet dari Twitter terkait PON XX dapat diolah dengan baik menggunakan metode *Naïve Bayes* untuk menganalisis sentimen positif dan negatif terkait acara tersebut [9].

Penelitian terdahulu yang berjudul “Penerapan *Text Mining* Analisis Sentimen Mengenai Vaksin Covid-19 Menggunakan Metode *Naïve Bayes*”, penelitian tersebut membahas tentang pembagian Vaksin Covid-19 di Indonesia dimana banyak masyarakat yang setuju dan banyak juga masyarakat yang tidak setuju akan vaksinasi ini karena dicap haram. Metode yang digunakan pada penelitian ini adalah *Naïve Bayes* dan data yang digunakan diambil dari *twitter* menggunakan API *twitter*, dan pengolahan datanya menggunakan *tools rapidminer*. Hasil dari analisis sentimen pada kata kunci vaksin covid-19 dengan menggunakan *confusion matrix* menggunakan algoritme *naïve bayes* [10].

Penelitian terdahulu yang berjudul “Penerapan Algoritme *Naïve Bayes* Untuk Menganalisis Sentimen *Review* Data *Twitter* BMKG Nasional”, membahas Perkembangan teknologi internet berdampak besar pada berbagai aspek kehidupan manusia, seperti pendidikan, bisnis, pemerintahan, dan komunikasi sosial. Survei pada tahun 2014 oleh Asosiasi Pengusaha Jasa Internet menunjukkan peningkatan jumlah pengguna internet di Indonesia, mencapai 20 juta pengguna pada tahun 2006. Media sosial, khususnya Twitter, menjadi platform yang populer di Indonesia dan digunakan oleh banyak orang untuk berinteraksi dan mengakses informasi. Opini dan keluhan masyarakat di Twitter dapat membentuk opini publik dan menjadi sumber evaluasi pelayanan yang diberikan oleh lembaga seperti BMKG. Penelitian ini menggunakan data sekunder dari akun Twitter resmi BMKG nasional dan mencapai akurasi sebesar 69,97% [11].

Penelitian terdahulu yang berjudul “Analisis Sentimen Jasa Transportasi *Online* pada Twitter Menggunakan Metode *Naïve Bayes Classifier*” membahas berkembangnya dunia teknologi informasi, penggunaan jasa transportasi online telah menjadi suatu kebutuhan bagi banyak orang. Oleh karena itu, analisis sentimen terhadap jasa transportasi online penting dilakukan untuk memahami tanggapan masyarakat terhadap layanan tersebut. Dalam penelitian ini, data yang digunakan berasal dari *platform* media sosial Twitter dan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan *tweet* menjadi kelas positif dan negatif. Hasil analisis menunjukkan bahwa persentase sentimen positif mencapai 88,60%, sedangkan sentimen negatif sebesar 11,40%, dengan tingkat akurasi mencapai 86,80%. Hal ini menunjukkan bahwa mayoritas *tweet* dari masyarakat memiliki sentimen positif terhadap jasa transportasi online [12].

Penelitian terdahulu yang berjudul “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritme *Naive Bayes Classifier*” yang membahas Twitter adalah platform media sosial yang populer, di mana publik dapat memberikan komentar bebas. Pemerintahan Joko Widodo sering menjadi target komentar publik, baik positif maupun negatif. Dalam penelitian ini, analisis sentimen dilakukan untuk mengklasifikasikan dokumen teks dari Twitter dalam Bahasa Indonesia menjadi sentimen negatif atau positif. Data terdiri dari 400 tweet, dengan 300 data latih dan 100 data uji. Menggunakan algoritme *Naïve Bayes Classifier*, penelitian ini mencapai akurasi 97%. Akurasi sentimen negatif adalah 96% dan sentimen positif adalah 98% [13].

Berdasarkan penelitian terdahulu yang berjudul "Implementasi *Text Mining* Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor" yang membahas Transportasi umum saat ini mengandalkan aplikasi berbasis internet, yang juga dikenal sebagai transportasi online. Grab dan Gojek adalah dua penyedia layanan transportasi online yang berusaha memberikan pelayanan yang baik kepada pengguna. Melalui platform Twitter, pengguna memberikan tanggapan, pengalaman, kritik, dan saran terhadap layanan yang diberikan oleh kedua perusahaan tersebut. Dalam penelitian ini, dikembangkan sebuah aplikasi untuk menganalisis opini masyarakat terhadap layanan Grab dan Gojek menggunakan metode text mining dengan algoritme *Naïve Bayes Classifier*. Hasil penelitian menunjukkan akurasi klasifikasi data uji sentimen untuk Grab sebesar 74,34%, sedangkan untuk Gojek sebesar 68,84%. Setiap kelas sentimen kemudian dianalisis untuk mengidentifikasi faktor-faktor yang mempengaruhinya. Pada Grab, ditemukan 6 faktor positif dan 5 faktor negatif, sedangkan pada Gojek ditemukan 8 faktor positif dan 6 faktor negatif. Setiap faktor kemudian diinterpretasikan dan divalidasi oleh pakar [14].

Berdasarkan penelitian sebelumnya, maka dilakukan penelitian untuk menganalisis sentimen media sosial twitter tentang digelarnya rapat KTT G20 di Indonesia menggunakan Metode Naive Bayes. Penelitian ini menggunakan metode pengambilan data melalui API Twitter dengan *tools netlytic* dengan pengelolaan 3.309 data *tweet*, untuk menghasilkan sentimen positif dan negatif dengan hasil yang divisualisasikan melalui gambar.

3. Metodologi

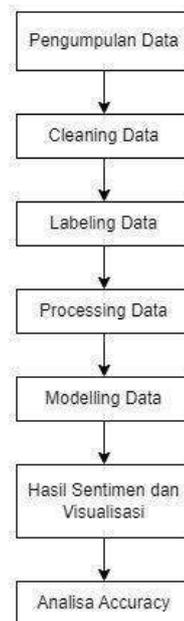
Metode *Naive Bayes* digunakan dalam penelitian ini karena metode ini cocok untuk menganalisis *dataset* emosional. Tujuan dari penelitian ini adalah untuk mendapatkan gambaran faktual dari pendapat pengguna Twitter. Bahan yang diperlukan untuk penelitian ini adalah diskusi di media sosial Twitter tentang KTT G20 di Indonesia. Data dikumpulkan menggunakan *netlytic*, yaitu perangkat lunak *platform* / situs web perangkat lunak berbasis *cloud* yang memiliki fungsi untuk pengambilan data, dan juga dapat digunakan untuk menganalisis teks dan grafik media sosial. Kemudian *tools* yang digunakan dalam penelitian ini adalah *Rapidminer*, yang merupakan *software open source* atau dapat dijelaskan bahwa *rapidminer* adalah *software* untuk menganalisis *data mining*, *text mining* dan analisa prediktif. *Rapidminer* mempunyai banyak sekali teknik pemrosesan untuk memberikan pemahaman yang baik kepada pengguna sehingga pengguna dapat membuat keputusan yang akurat, dibangun di sekitar 500 operator data, termasuk operator untuk entri data, pemrosesan awal, dan ringkasan hasil. *Rapidminer* bisa digunakan di semua jenis OS baik itu *windows* atau *linux* [15].

Waktu penarikan data sumber dari *netlytic*, dijelaskan proses penyederhanaan data, dengan proses memilah data dengan kata *keyword* yang sedang dicari, kemudian pemaparan data, pemaparan data penting untuk lebih mudah dipahami, yang nantinya bisa divisualisasikan dalam bentuk gambar, tabel, atau skema. Penelitian ini menggunakan data yang sudah diambil di *netlytic* sebagai objek data uji coba, kemudian data objek yang sudah dilatih, akan berguna sebagai data uji untuk data yang belum memiliki sentimen, bisa dibilang data latih yang digunakan adalah data *training* yang digunakan *naïve naves* melakukan klasifikasi.

Untuk penarikan kesimpulan, data yang didapat akan dianalisa, dan bisa membantu orang-orang untuk lebih bisa mengira atau menebak bagaimana respon orang terhadap suatu kejadian yang hanya berdasarkan data *tweet* saja. Penelitian dilaksanakan melalui beberapa proses, antara lain: pengumpulan data, *cleaning* data, *labeling* data, *processing* data, *modelling* data, dan analisa dari hasil yang didapat. Penjelasan untuk proses proses penelitian ini akan dijelaskan sebagai berikut.

Gambar 1 menunjukkan tahapan penelitian yang terdiri dari beberapa tahapan yang dijelaskan sebagai berikut, dimulai dengan fase registrasi, lalu fase pengeditan dengan membuat proyek *new data*. Pada proses ini, anda dapat klik tombol "*new dataset*" untuk mengambil data baru dan memasukkan kata kunci hastag G20. Langkah selanjutnya adalah pratinjau. Pada langkah selanjutnya, *netlytic* akan menyajikan hasil pencarian yang disajikan di *spreadsheet* dalam format *Excel* dan *CSV*. Berkas pratinjau biasanya tersedia untuk diunduh. Langkah selanjutnya adalah melakukan *cleaning* data, pada bagian data *excel* akan di-*import* ke *rapidminer* untuk *cleaning* data, pada tahap ini data yang memiliki *hashtag*, *mention*, *link*, atau bahkan semua tanda baca dan data yang *duplicate* akan dihilangkan yang bertujuan untuk mempermudah tahap *Processing* data. Tahap berikutnya adalah *labelling* data, data yang sudah

dibersihkan tadi akan diproses *labelling* data yang nantinya akan digunakan oleh *Naïve Bayes* sebagai data latih. Kemudian masuk pada tahap pengolahan data, tahap ini melakukan menyaring data berdasarkan data sentimen yang ada, selanjutnya tahap merubah data nominal menjadi teks.



Gambar 1 Tahapan penelitian

Kemudian melakukan *processing* data untuk menentukan seberapa sering muncul kata-kata dari data yang ada, selanjutnya membuat *role set* menggunakan data sentimen, *role set* merupakan identitas dan nama atribut dari *text*, kemudian data yang *ready* dan sudah dilabeli sentimen positif dan negatif akan masuk ke *process learning* menggunakan metode *Naïve Bayes*, algoritme *Naive Bayes* menggunakan perhitungan probabilitas untuk menyederhanakan data yang kompleks menjadi probabilitas-probabilitas yang lebih sederhana. Metode ini membutuhkan data pelatihan untuk menentukan parameter-parameter yang diperlukan dalam proses klasifikasi. Untuk formula dari *Naïve Bayes* sebagai berikut

$$P(X) = \frac{p(X|H)p(H)}{p(x)} \dots\dots\dots (1)$$

Keterangan:

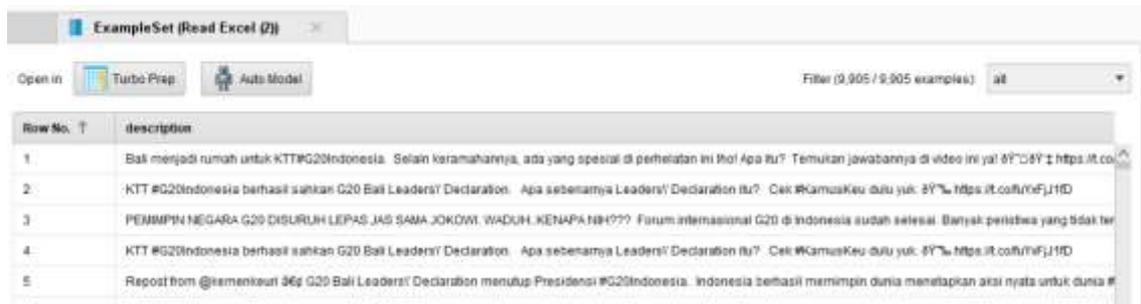
- X = Data kelas yang belum diketahui
- H = Hipotesa data X adalah suatu kelas spesifik
- P(H|X) = Probabilitas hipotesis H pada kondisi X (*posterior probability*)
- P(H) = Probabilitas hipotesis H (*prior probability*)

Agar *naïve bayes* mengerti sentimen negatif dan positif yang ada didalam data tersebut. Data yang sudah diajari oleh *naïve bayes* akan ditempatkan di operator *store model* yang berfungsi untuk menyimpan data *model* nya, serta data uji coba akan ditempatkan di operator *store* latih, data-data tersebut akan disimpan yang nanti akan digunakan untuk tahap berikutnya dimana tahap selanjutnya adalah proses *Union* data yang sudah diajari *naïve bayes*. Selanjutnya masuk ke tahap *modelling* data, tahap ini adalah tahap penggabungan data yang sudah diajari tadi dengan data uji coba, proses ini akan mengisi data uji coba secara otomatis yang akan dilakukan oleh *naïve bayes*. Kemudian setelah selesai itu data uji coba akan terisi secara otomatis dan sudah ada sentimen positif dan negatifnya, selanjutnya dari data yang sudah selesai sudah bisa dilakukan analisis sentimen tentang G20.

4. Hasil dan Pembahasan

Pada penelitian ini menggunakan data sebanyak 9.905 data set yang berupa data *tweet*, pada penarikan data menggunakan *netlytic* dengan *keyword* atau *hashtag* G20 sebagai objek

penelitian ini, setelah data di export menjadi excel yang dimasukkan ke dalam *RapidMiner* untuk masuk tahap yang pertama. Untuk lebih jelas bisa dilihat pada Gambar 2.

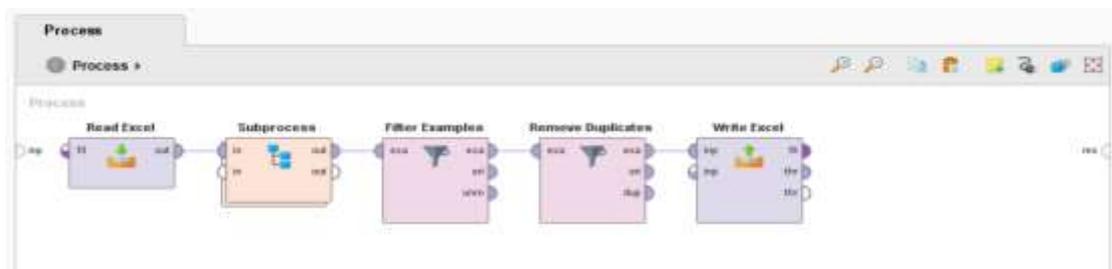


Gambar 2 Tampilan data mentah (Source: Rapidminer)

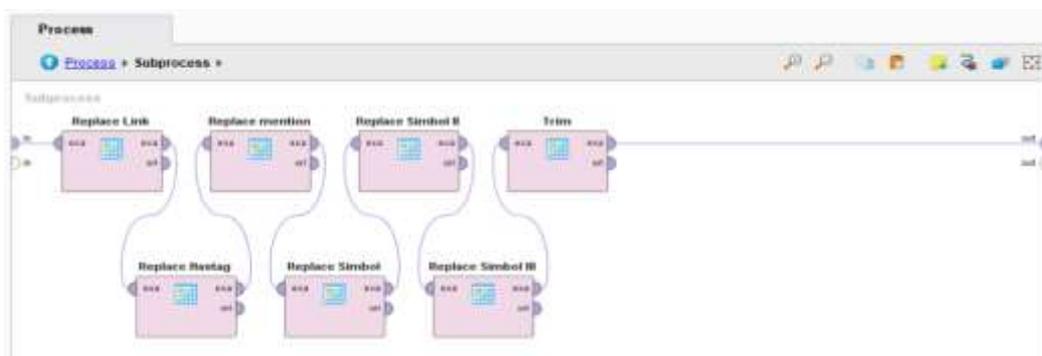
4.1. Pemrosesan Data

1) **Cleaning Data**

Pada proses *cleaning* data yang diambil menggunakan *netlytic* adalah 9905 data, kemudian data tersebut masuk ke tahap *cleaning* dengan menyisakan data yang sudah bersih sebanyak 3309 data. Proses pada tahap ini menggunakan beberapa operator di antaranya; operator *Read Excel* yang berguna untuk memasukkan data ke *rapidminer*, kemudian ada operator *Subprocess* yang di dalamnya ada beberapa proses lagi, selanjutnya ada operator *Filter Examples* yang berfungsi untuk menghilangkan data yang kosong, kemudian ada operator *Remove Duplicates* operator ini berfungsi menghilangkan data yang sama, dan terakhir ada operator *Write Excel* berfungsi untuk menyimpan data yang sudah dibersihkan oleh operator sebelumnya, Untuk lebih jelas proses bisa dilihat pada Gambar 3.



Gambar 3 Proses Pembersihan Data (Source: Rapidminer)



Gambar 4 Isi Operator Subprocess (Source: Rapidminer)

Kemudian masuk kedalam bagian subproses yang ada pada *cleaning* Data subproses berguna untuk menyimpan beberapa proses yang terlalu banyak agar lebih mudah untuk pengolahan data, di dalam operator *Subprocess* ini terdapat beberapa operator yang berfungsi

untuk *replace* semua simbol atau karakter unik yang ada di dalam data *tweet*, Untuk lebih jelas bisa dilihat pada Gambar 4.

2) Labeling Data

Kemudian masuk ke tahap pelabelan data uji, pada tahap ini data dilabel dengan cara manual dan data yang digunakan sebanyak 700 data dari 3.309 data yang sudah dibersihkan sebelumnya, dan pada saat pelabelan secara manual kami memberi sentimen positif dan negatif berdasarkan isi *tweet* yang ada, sebelum masuk proses *processing* dan proses pelatihan oleh *naïve bayes*. Untuk lebih jelas bisa dilihat pada Gambar 5.

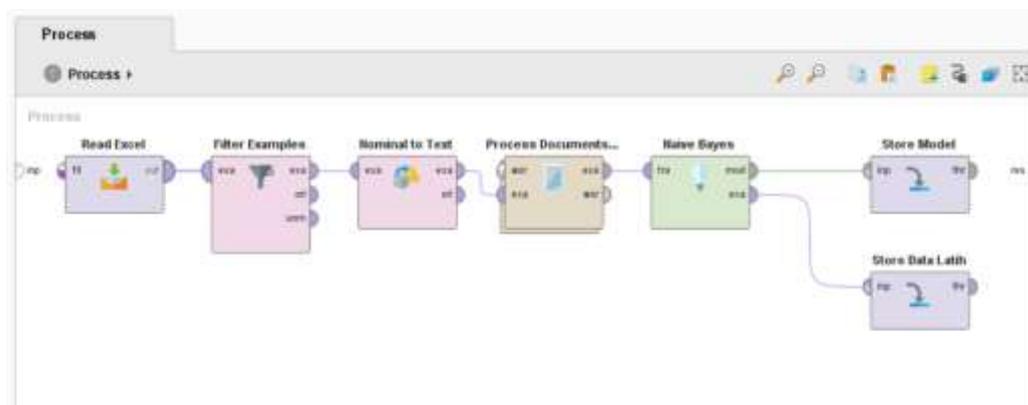


Row No.	Sentimen	Description
1	positif	Baik menjadi rumah untuk KTT Selain keramahannya ada yang spesial di perhelatan ini lho Apa itu? Tentukan jawabannya di video ini ya
2	positif	KTT berhasil sahkan Gwerty Bali Leaders! Declaration: Apa sebenarnya Leaders! Declaration itu? Cek dulu yuk!
3	negatif	PEMIMPIN NEGARA Gwerty DIGURUH LEPAS JAS SAMA JOKOWI WADUKENAPANNI Forum internasional Gwerty di Indonesia sudah selesai Banyak pembicaraan
4	positif	Repositif from Gwerty Bali Leaders! Declaration menup Presiden! Indonesia berhasil memimpin dunia menetapkan aksi nyata untuk dunia dan mendorong pr
5	positif	Penyelenggaraan KTT Gwerty di Indonesia nyata memberikan banyak dampak positif salah satunya kepercayaan investor pada iklim bisnis yang baik di Indonesia
6	positif	KTT Gwerty menjadi puncak rangkaian Presiden! Gwerty Indonesia selama setahun penuh. Berbagai aksi nyata dan terobosan besar di jalur keuangan untuk
7	positif	KTT Gwerty berikan dampak positif pada perekonomian Indonesia khususnya maupun sektor UMKM
8	positif	Disarik suksesnya perhelatan antar KTT Gwerty ada banyak momen yang dapat menarik mata warganet salah satunya momen para Pemimpin Dunia Penasara

Gambar 5 Contoh Data yang Sudah dilabeling (Source: Rapidminer)

3) Processing Data

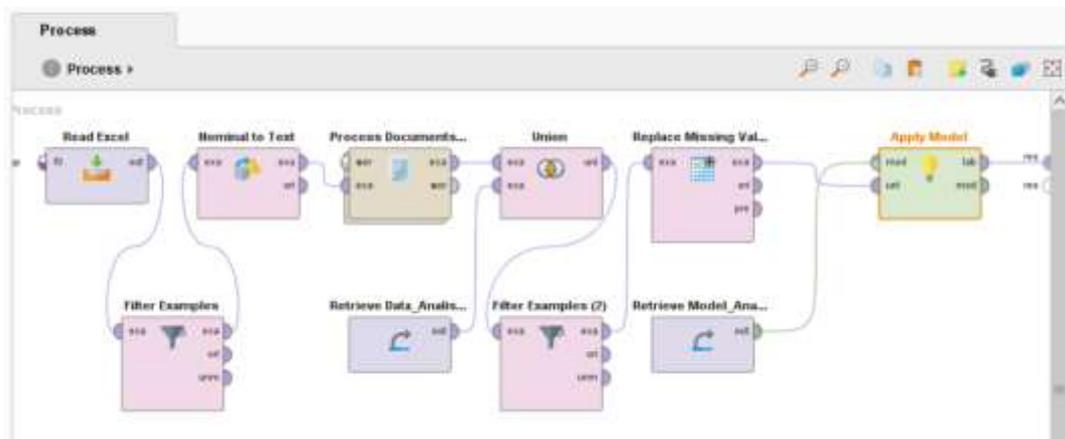
Kemudian masuk ke *Processing Data*, data yang sudah diinput akan masuk ke tahap *filter examples* didalamnya ada parameter *filter "sentiment is not missing"*, tahap ini berfungsi untuk memilah file yang masuk tidak *null* pada bagian kolom sentimen, masuk ke proses selanjutnya, ialah operator "*nominal to text*". Kemudian masuk operator *process document from data*, operator ini berfungsi untuk tahap *processing data*, dalam proses ini ada sub operator didalamnya ada *tokenize* menggunakan format *non letters* yang memilah antar kata di data yang ada, yang kedua operator *transform cases* dan *transform to lowercase* berguna untuk mengubah huruf menjadi kecil, selanjutnya *Filter Tokens (by Length)* biasanya menggunakan parameter *max chars 25* dan *filter min chars 2*, jika data memiliki lebih dari dari 25 karakter dan kurang dari 2 karakter akan hilang. Pada langkah selanjutnya, operator *naïve bayes* digunakan untuk mengklasifikasikan catatan berlabel data (700 data) untuk mendapatkan model data yang dilatih oleh operator *naïve bayes* dan digunakan sebagai *library* untuk pelabelan otomatis. Kemudian gunakan operator *store* yang berfungsi untuk menyimpan data model dan operator *store* lainnya untuk menyimpan data latih yang disimpan di *rapidminer*. Setelah tersimpan semua, pembelajaran *naïve bayes* akan masuk ke pelabelan otomatis. Untuk proses ini dapat dilihat pada Gambar 6.



Gambar 6 Proses Labeling Data (Source: Rapidminer)

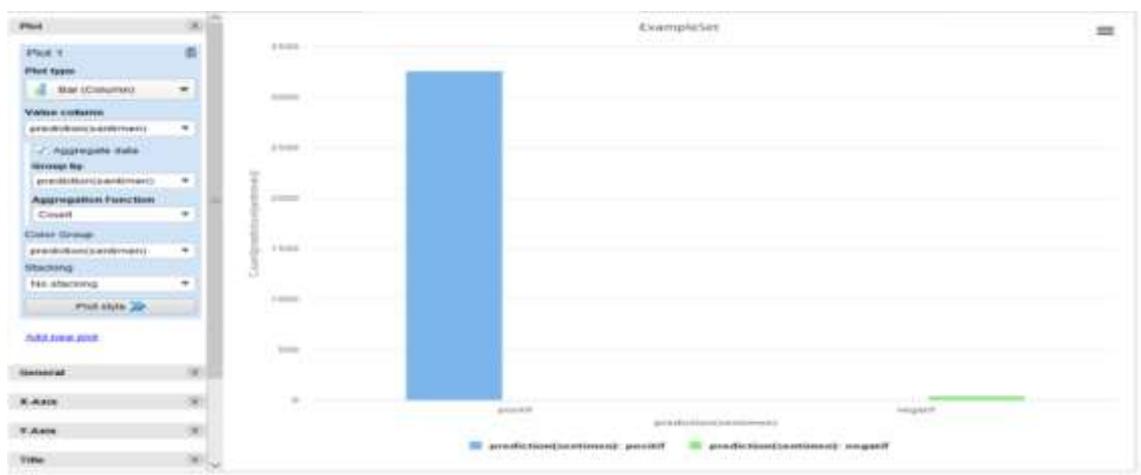
4) Modeling Data

Pemodelan data uji adalah langkah yang secara otomatis melabeli data kosong yang belum ada isi label sentimen positif atau negatif, proses pertama adalah menggunakan operator *read excel* yang berfungsi untuk memasukkan data ke *rapidminer* dan kemudian memfilter data dengan parameter “*sentiment is missing*”, hal ini dilakukan untuk memfilter kumpulan data sehingga data yang diinputkan hanyalah informasi kosong, kemudian bagian deskripsi difilter menggunakan parameter *is not missing*, maka langkah berikutnya adalah operator “*Nominal to Text*”, kemudian “*Process Document From Data*” yang berfungsi sebagai operator yang memproses *dataset* dengan 4 sub-operator, diantaranya “*tokenize*” dengan patokan *non-letters*, berfungsi memisahkan kata di dalam *tweet*, selanjutnya operator “*transform cases*” dan “*transform to lowercase*” berfungsi mengubah semua abjad menjadi huruf kecil. Selanjutnya operator filter “*Tokens (by Length)*” dengan parameter filter minimal 2 karakter dan maksimal 25 karakter. Jika kata < 2 dan > 25, kata tersebut akan dihilang. Selanjutnya masuk ke operator *Union*, yaitu operator yang menggabungkan dua tabel menjadi baris baru dengan menggabungkan data yang sama dengan class dan data training, setelah digabung didapatkan *naïve bayes machine learning*. dilanjutkan dengan operator “*replace missing values*” berguna *replace* kolom yang kosong ke kolom yang sudah di-set 0, selanjutnya masuk ke operator “*apply model*” berfungsi untuk menggunakan model yang sudah disimpan sebelumnya, dengan begitu *naïve bayes* dapat memberikan data kolom sentimen positif negatif, menurut model yang dibuat *naïve bayes*, untuk prosesnya dapat dilihat pada Gambar 7.



Gambar 7 Modeling Data Training (Source: Rapidminer)

4.2. Hasil dan Visualisasi



Gambar 8 Visualisasi Sentimen Positif dan Negatif (Source: Rapidminer)

Langkah selanjutnya menggunakan tipe visualisasi *bar (Column)*, dimana nilai kolom = prediksi (sentimen), kemudian memeriksa data agregat dan untuk grup = prediksi (sentimen), selanjutnya untuk *aggregation function = count*, lalu opsi *sorting = value column, direction = descending*, selanjutnya opsi kelompok warna = prediksi, untuk *stacking = no stacking*, terakhir gaya gambar yang digunakan adalah menggunakan settingan bawaan. Pada layar visualisasi, ada bar berwarna biru menunjukkan prediksi opini (sentimen) positif: Positif = 3.264 dan hijau adalah perasaan negatif dengan perasaan: Negatif = 45, sehingga jumlah data yang dimasukkan dalam analisis opini adalah 3.309. Untuk hasil proses dapat dilihat pada Gambar 8.

Analisis akurasi berdasarkan model metode *naïve bayes* dapat dilakukan dengan menggunakan 700 data yang sudah di label, 675 sentimen positif dengan *class precision* 100 %, *class recall* 99,56%, 22 sentimen negatif dengan *class precision* 88,00%, *class recall* 100,00%, dengan presisi kelas dan perhitungan presisi secara manual sebagai berikut:

$$Accuracy = \frac{(pred\ positif\ true\ positif + pred\ negatif\ true\ negatif) \times 100}{Data\ Latih}$$

$$Accuracy = \frac{(675 + 22) \times 100}{700}$$

$$Accuracy = 99.57142857 \%$$

Dan perhitungan otomatis pada *rapidminer* adalah 99.57%, dan dapat dilihat pada Gambar 9.



	true positif	true negatif	class precision
pred. positif	675	0	100.00%
pred. negatif	3	22	88.00%
class recall	99.56%	100.00%	

Gambar 9 Analisa Accuracy (Source: Rapidminer)

Pada gambar diatas bisa dilihat hasil pengujian akurasi algoritme Naïve Bayes mencapai 99.57%. Selain itu, nilai precision untuk pred.positif mencapai 100%, nilai precision untuk pred.negatif mencapai 88% dan recall untuk true positif mencapai 99.56%, nilai recall untuk true negatif mencapai 100%. Temuan ini menunjukkan bahwa algoritme Naïve Bayes adalah pilihan yang baik dalam menganalisis sentimen pengguna media sosial terhadap isu-isu yang dibahas di G20.

5. Simpulan

Berdasarkan hasil pengolahan data dan pembahasan analisis sentimen masyarakat terhadap data *tweet* G20 menggunakan metode *naïve bayes*, hasil nilai analisa *accuracy* yang didapat adalah *prediction* negatif 45 dan *prediction* positif 3.264. Berdasarkan output yang didapat dari data dapat diambil kesimpulan bahwa dari *tweet* di media sosial Twitter lebih banyak yang berpendapat positif untuk KTT G20 yang diselenggarakan di Indonesia, Indonesia dan KTT G20 ini bisa membantu pemerintah untuk menjadikan acuan pengambilan keputusan dalam menyelenggarakan kegiatan kenegaraan berikutnya dan membantu pemerintah untuk menjelaskan kepada masyarakat bahwa penyelenggaraan G20 di Indonesia bertujuan untuk menyejahterakan masyarakat dan menjadi salah satu cara untuk mendorong perekonomian negara ini agar lebih baik lagi dan. Saran untuk penelitian selanjutnya yaitu menggunakan metode *support vector machine* untuk bisa mendapatkan perhitungan *accuracy* lebih baik lagi.

Daftar Pustaka

- [1] Bank Indonesia, "Presidensi G20 Indonesia 2022," 2022. <https://www.bi.go.id/id/g20/default.aspx> (diakses 10 Mei 2023).
- [2] Kemenkeu.go.id, *G20pedia - Informasi Presidensi G20 Indonesia 2022*. 2022. Diakses: 23 Mei 2023. [Daring]. Tersedia pada: <https://anggaran.kemenkeu.go.id/api/Medias/53830b3e-a257-4a06-ae33-681a404eace7>
- [3] A. Fransisca Putri, I. Ketut Merta, dan I. Ayu Sasmitha Putri, "Putri, Merta, Sasmitha Putri Pengaruh Implementasi Leadership Indonesia Terhadap Presidensi G20 Dan Pertumbuhan Ekonomi Di Bali," vol. 8, no. 2, 2022, doi: 10.47329/jurnal_mbe.v8i2.894.
- [4] M. Kartika, S. Saepudin, dan D. Gustian, "Analisis Sentimen Dampak Covid-19 Terhadap Pembatalan Keberangkatan Ibadah Haji Pada Tahun 2020," 2021.
- [5] N. Hannani, "Pengertian Twitter Beserta Sejarah dan Manfaat Twitter (Lengkap)," *Nesabamedia*, 6 Oktober 2019. <https://www.nesabamedia.com/pengertian-twitter/> (diakses 17 November 2022).
- [6] A. H. Anshor, "Analisis Sentimen Warganet Terhadap KTT G20 Bali Menggunakan Algoritme Naïve Bayes," *Jutisi*, vol. 11, 2022.
- [7] M. K. Sandryan, B. Rahayudi, dan D. E. Ratnawati, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritme Backpropagation dan Term Frequency-Inverse Document Frequency," 2021. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [8] S. Fendyputra Pratama, R. Andrean, dan A. Nugroho, "Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 4, no. 2, hlm. 2541–3619, 2019, doi: 10.31328/jo.
- [9] K. Jonathan Sitorus dan A. Muliawati, "Klasifikasi Dan Analisis Sentimen Pada Data Twitter Menggunakan Algoritme Naïve Bayes. (Studi Kasus: Pekan Olahraga Nasional XX 2021)," 2022. [Daring]. Tersedia pada: <https://t.co/dsH2pZdXvd>
- [10] F. Fathonah dan A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid -19 Menggunakan Metode Naïve Bayes," *Jurnal Sains dan Informatika*, vol. 7, no. Volume 7 Nomor 2, November 2021, 2021.
- [11] D. Darwis, N. Siskawati, dan Z. Abidin, "Penerapan Algoritme Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal TEKNO KOMPAK*, vol. 15, no. 1, hlm. 131–145, 2021.
- [12] B. Mas Pintoko dan K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," Bandung, Des 2018.
- [13] Y. S. Mahardhika dan E. Zuliarso, *Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritme Naives Bayes Classifier*. 2018.
- [14] I. Olive, D. Putra, K. Rega Prilianti, P. Lucky, dan T. Irawan, "Implementasi Text Mining Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor," vol. 8, no. 2, 2020.
- [15] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, dan C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, hlm. 8–17, Mei 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.