

Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter Terhadap KTT G20 di Indonesia

Herda Andriana¹, Shofa Shofia Hilabi^{2*}, Agustia Hananto³

Fakultas Ilmu Komputer, Sistem Informasi, Universitas Buana Perjuangan Karawang, Karawang, Indonesia

Email: ¹si19.herdaandriana@mhs.ubpkarawang.ac.id, ^{2*}shofa.hilabi@ubpkarawang.ac.id, ³agus.tia@ubpkarawang.ac.id

Email Penulis Korespondensi: shofa.hilabi@ubpkarawang.ac.id

Submitted 29-12-2023; Accepted 10-02-2023; Published 17-02-2023

Abstrak

Baru-baru ini Indonesia menjadi tuan rumah dalam acara besar yang diadakan di Bali pada 15 November 2022 yaitu pada KTT (Konferensi Tingkat Tinggi) G20. Tujuan didirikannya G20 tidak lain yakni meningkatkan perekonomian global yang sempat krisis. Tetapi keikutsertaan Negara Indonesia menjadi anggota KTT G20 mengundang kontroversial sehingga menimbulkan berbagai pendapat dari masyarakat khususnya masyarakat Negara Indonesia itu sendiri sehingga dibutuhkan analisis sentimen terhadap KTT G20. Analisis sentimen di ambil dari tweet mengenai KTT G20 yang ditulis pada sosial media twitter. Pengumpulan data yang berupa tweet dengan cara scraping mendapatkan data sebanyak 2500 tweet. Dari banyaknya data yang telah di kumpulkan maka dibutuhkan metode untuk mengklasifikasi tweet termasuk kelompok positif, netral atau negatif. Tujuan dilakukannya penelitian ini adalah untuk analisis opini publik di Twitter pada KTT G20. Pengolahannya menggunakan perangkat lunak Orange dengan beberapa tools yang digunakan dan metode K-Nearest Neighbor sehingga mendapatkan hasil berupa 1107 data tweet masuk pada kelompok original dengan tingkat accuracy, F1 score, Recall, Precision rata-rata 99%. Dan hasil sentimen analisis 89 tweet masuk kedalam kategori negatif, 614 netral, dan 404 positif dengan mayoritas tingkat emosional joy, surprise, dan fear.

Kata Kunci: K-Nearest Neighbor; KTT G20; Twitter; Sentimen Analisis; Orange

Abstract

Indonesia will host the KTT (Konferensi Tingkat Tinggi) G20 summit on the island of Bali on November 15, 2022. The G20 was formed with one goal in mind: to boost the global economy, which had just entered a period of crisis. However, Indonesia's participation as a full member of the Group of Twenty (G20) has sparked controversy among the country's general populace and the population of Indonesia itself, necessitating a thoughtful analysis of the group's motives. Sentiment analysis was gleaned from tweets on KTT G20 posted on the social media platform Twitter. Data scraping yielded a total of 2,500 tweets for inclusion in the collection. Methods for classifying tweets into positive, neutral, and negative groups are required because of the large amount of data that has already been collected. The purpose of this study was to analyze public opinion on Twitter during the KTT G20. The data was processed using the Orange neural network using a number of tools and the K-Nearest Neighbor method, yielding a total of 1,107 tweets that were successfully added to the original set, with an average recall and precision of 99%. According to the analysis of sentiment, there were 89 negative tweets, 614 neutral tweets, and 404 positive tweets, with the most common emotions being happiness, surprise, and fear.

Keywords: K-Nearest Neighbor; KTT G20; Twitter; Sentiment Analysis; Orange

1. PENDAHULUAN

Krisis keuangan global tahun 2007 berdampak negatif baik bagi negara maju maupun negara berkembang karena dampak kemajuan teknologi dan liberalisasi perdagangan internasional. Akar penyebab krisis keuangan internasional dapat ditelusuri kembali ke kurangnya regulasi sektor keuangan yang kritis dan manajemen krisis global. Krisis ekonomi global telah menyebabkan jatuhnya sektor keuangan di negara maju dan mempengaruhi sektor riil, sedangkan negara berkembang berdampak pada perdagangan dan investasi luar negeri. Peristiwa ini memunculkan ide G20 (*Leader Meeting Group*) untuk menyelenggarakan KTT G20 bagi para kepala negara atau pemerintahan yang berfokus pada perekonomian dunia yang ditetapkan pada tanggal 15 November 2008 di Washington[1]. KTT G20 (*Group of Twenty*) merupakan kelompok beranggotakan 20 Negara merupakan terdiri dari 19 Negara dan satu lembaga Uni Eropa yakni Negara Australia, Argentina, Brazil, Kanada, China, Uni Eropa, Jerman, Prancis, India, Indonesia, Meksiko, Jepang, Italia, Saudi Arabia, Rusia, Afrika Selatan, Korea Selatan, Turki, Inggris Raya, Dan Amerika Serikat[2].

Pada November 2022, Indonesia menjadi tuan rumah KTT G20 yang dilaksanakan di Bali. Menurut Menteri Keuangan Negara Indonesia, G20 merupakan forum penting untuk kerja sama ekonomi internasional dengan ekonomi terbesar di dunia[3]. Hal ini cukup strategis karena G20 menyumbang lebih dari 60% populasi dunia, 75% perdagangan dunia, dan 80% PDB dunia[4]. Keikutsertaan Indonesia dalam KTT G20 menandai upaya pertama negara tersebut untuk memastikan produknya memenuhi standar kualitas internasional dan menarik bagi konsumen di seluruh dunia. Kedua, Indonesia akan memiliki akses ke peluang pasar yang lebih luas berkat kemajuan teknologi dan kolaborasi. Penting bagi Indonesia, yang berperan penting dalam menjaga stabilitas kawasan, untuk ikut serta dalam KTT G20[5]. Tetapi hal ini sempat menjadi kontroversial yang menyebabkan munculnya berbagai opini warganet khususnya masyarakat Indonesia yang dipaparkan melalui postingan, komentar dan tanggapan di media sosial salah satunya yakni Twitter[6].

Twitter banyak digunakan masyarakat Indonesia sebesar 59% dan merupakan media sosial ke-5 yang paling banyak digunakan pada tahun 2020[7][8]. Menyampaikan pendapat di twitter bisa menjadi salah satu media yang dijadikan penelitian[9]. Dari twitter tentunya dapat menemukan banyak opini publik tentang KTT G20 yang terdiri dari komentar positif dan negatif serta menimbulkan kekhawatiran tersendiri. Dengan demikian, solusi mungkin diperlukan untuk mengklasifikasikan komentar sebagai positif, netral, atau negatif menurut metode *K-Nearest Neighbor* untuk

menganalisis sentimen publik tentang KTT G20. K-Nearest Neighbor adalah metode untuk mengklasifikasikan target berdasarkan data training dengan menghitung jarak dari target ke tetangganya. Ketika sampai pada fase pelatihan, algoritma diterapkan untuk menangani hal-hal seperti mendaftarkan fitur dalam vektor pelatihan dan mengklasifikasikan data pelatihan. Keuntungan dari metode K-Nearest Neighbor adalah rekam jejak algoritma klasifikasi yang terbukti menghasilkan hasil yang akurat dan konsisten terlepas dari platform komputasi yang digunakan dalam tugas klasifikasi yang ada[10][11].

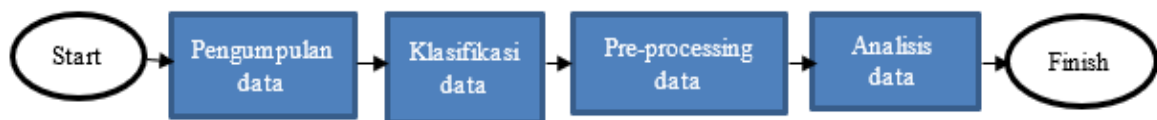
Beberapa penelitian sebelumnya menggunakan metode K-Nearest Neighbor terhadap analisis sentimen. Penelitian dengan judul “Analisis Sentimen Pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya Dengan Metode K-Nearest Neighbor” yang dilakukan oleh Luqman dan kawan-kawan menyimpulkan hasil berupa mendapatkan tingkat akurasi sebesar 86%[12]. Penelitian yang berjudul “Implementasi K-Nearest Neighbor Dalam Analisis Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter” dilakukan oleh Rasenda dkk, dengan hasil membuktikan nilai accuracy 70,59%, precision 69,87%, dan recall 62,32%[13]. Penelitian yang dilakukan oleh Faiza dkk, berjudul “Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbor” dengan hasil dari penelitian yang memiliki kesimpulan akurasi 79,43% dengan sentimen positif 47,59%, negatif 8,66% dan netral 43, 75%[14]. Penelitian berjudul “Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor” yang dilakukan oleh arif dan Herny dengan hasil yang di dapatkan akurasi 69.5 persen, recall 69.5 persen, dan presisi 68.5 persen[15]. Penelitian dengan judul “K-Nearest Neighbor Untuk Menganalisis Sentimen Terhadap Kebijakan Merdeka Belajar Kampus Merdeka Pada Komentar Twitter” yang dilakukan oleh Rahmad dkk dengan hasil akurasi maksimum sebesar 84,28%, precision 84,22%, recall 84,28% dan f-measure 84,25%[16]. Penelitian yang dilakukan tentunya berbeda dengan penelitian sebelumnya yaitu berbeda dalam software pengolahan data dan topik. Dan memiliki persamaan yaitu dalam menggunakan metode. Penelitian ini memiliki tujuan untuk menganalisis sentimen masyarakat twitter menggunakan metode K-Nearest Neighbor melalui empat proses utama yaitu *scraping*, klasifikasi, *word processing*, kemudian analisis dengan *bantuan software orange* yang diterapkan dalam penelitian ini. Hasil klasifikasi tersebut selanjutnya positif, negatif atau netral dan nantinya akan digunakan sebagai ukuran seberapa baik atau buruk pendapat masyarakat Indonesia terhadap KTT G20 tersebut.

2. METODOLOGI PENELITIAN

Isu seputar pelaksanaan KTT G20 di Bali menjadi objek penelitian yang dianalisis dan dibahas dalam penelitian ini. Algoritma K-Nearest Neighbor digunakan untuk mengklasifikasikan sikap komunitas sebagai positif atau negatif. Pemrosesan dan analisis data dengan rangkaian program Orange data mining.

2.1 Tahapan Penelitian

Pengujian ini menggunakan perangkat lunak *Orange* untuk analisis dan akan dilanjutkan pada fase-fase berikutnya :



Gambar 1. Alur Penelitian.

a. Pengumpulan Data (*scraping*)

Penelitian ini menggunakan data sebanyak 2500 yang di peroleh dari twitter dengan cara *scraping* menggunakan beberapa hastag yang berkaitan dan sesuai dengan topik penelitian, yakni #KTTG20 #G20 pada *website* Netlitic.org. Data yang digunakan dari tanggal 15 November 2022 hingga 7 hari kebelakang.

b. Klasifikasi Data

Klasifikasi data pada penelitian ini menggunakan metode *K-Nearest Neighbor* (KNN), merupakan salah satu metode klasifikasi non parametik. Prinsip operasinya adalah mengklasifikasikan objek pada tes set berdasarkan mayoritas kelas dari k tetangga terdekat pada training set. Nilai K adalah bilangan bulat positif yang ditentukan oleh pengguna. Perhatikan bahwa semakin kecil nilai k, semakin sensitif terhadap efek noise/runway, sedangkan semakin besar nilai k, maka semakin kecil efek noise/runway, tetapi batas antar kelas menjadi kabur. Untuk mencari nilai K yang optimal dapat digunakan teknis *cross-validation*[6][17]. Jarak euclidean dapat dihitung dengan persamaan berikut[18] :

$$d(p,q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Pada tahapan klasifikasi data yang telah didapatkan dengan cara *scraping* sebanyak 2500 kemudian di klasifikasi kedalam beberapa kelompok, seperti tweet original, retweet, quotes, atau reply untuk mengetahui dari 2500 data berapa banyak yang masuk kedalam masing-masing kelompok.

c. Pre-processing Data

Preprocessing digunakan untuk menyaring data mentah yang kemudian dipersiapkan untuk diolah nantinya. Tujuan dari tahap pengolahan data adalah untuk menyaring data mentah yang diperoleh pada proses *scraping* dan mempersiapkannya untuk tahap berikutnya. Fase *preprocessing* data terdiri dari tiga fase[19]:

1. Transformation

- Lower case* berfungsi mengubah data awal menjadi huruf kecil. Implementasi pada kasus ini terdapat pada contoh kata 'Lokasi' menjadi 'lokasi' dan 'SENYUM' menjadi 'senyum'. Kemudian pada kata 'PulihBersama' menjadi 'pulihbersama'.
- Remove accents* berfungsi untuk menghilangkan diakrtitik. Contohnya terdapat pada kata 'rècover' menjadi 'recover'.
- Parse HTML* berfungsi mengenali tag HTML dan mengubahnya menjadi kalimat biasa.
- Remove URLS* berfungsi untuk menghilangkan url. Pada tahapan ini membantu untuk menghilangkan kata atau kalimat yang mengandung url, seperti <https://www.dll>.

2. Tokenization

Regex digunakan untuk mengekstraksi teks berdasarkan ekspresi reguler di perangkat lunak *Orange*.

3. Filtering

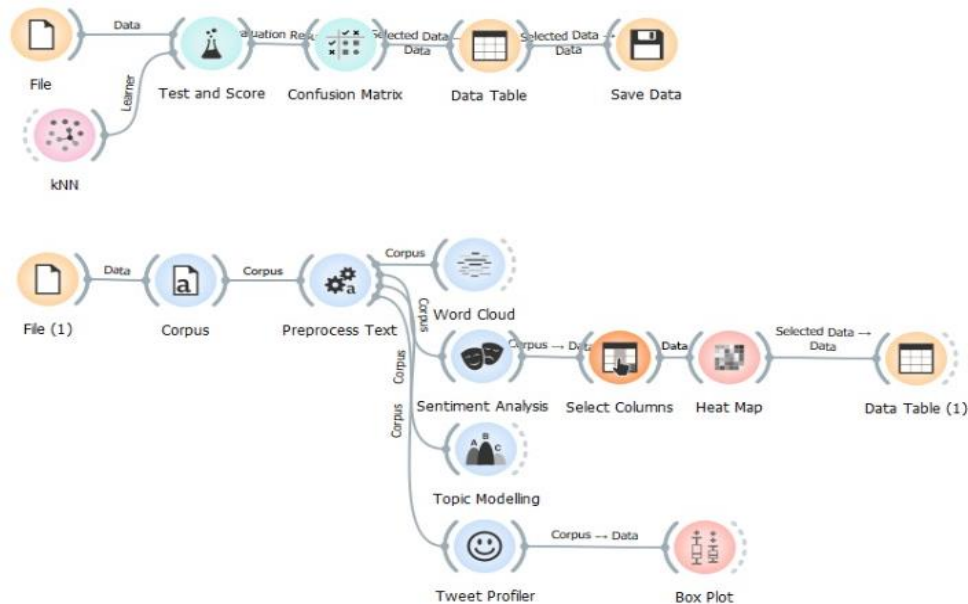
- Stopword* berfungsi memberhentikan kata dari teks. Aplikasi data mining *Orange* juga memiliki library bahasa Indonesia, atau bisa mengetikkan kata yang akan dihilangkan tentunya dibuat sendiri dengan file format .txt
- Regex Symbols* berfungsi untuk menghapus tanda baca seperti contoh (`|,:|;|!|?|\(|\)|\\||+|'|\"'|"'"|...|-|--|_|$&|*|>|<|)`).

d. Analisis Data

1. *Word cloud* adalah metode menggunakan teks, menampilkan grafik frekuensi kata yang sering muncul dalam teks sumber. Semakin besar ukuran kata dalam gambar, semakin sering muncul dalam data penelitian[20]. Pada pembobotan kata ditahap ini kata yang sering muncul pada dokumen yaitu kata G20 sebanyak 2703 dan merupakan kata yang paling besar ukurannya dibanding kata lain.
2. *Topic Modeling* secara konseptual, sebuah topik terdiri dari kata-kata yang mendefinisikannya, dan satu dokumen dapat berisi beberapa topik berbeda, masing-masing dengan peluang uniknya sendiri. Ini adalah ide inti dibalik pemodelan topik. Namun, dari sudut pandang manusia, dokumen adalah objek berwujud yang dapat di pahami; topik, distribusi topik dalam dokumen, dan pengelompokan kata di dalam topik adalah struktur abstrak. Inilah sebabnya mengapa pemodelan topik adalah untuk mengidentifikasi topik dan kata-kata yang menyusunnya[21]. Pada tahapan ini ditemukan beberapa topik yang diambil berdasarkan 10 terbanyak yang akan ditampilkan pada pembahasan.
3. *Sentiment Analysis* Secara umum, terbagi dalam dua kategori umum:
Coarse-grained sentiment analysis : melakukan proses analisis pada tingkat dokumen dan merupakan salah satu dari dua kategori besar yang digunakan untuk mengklasifikasikan analisis sentimen. Singkatnya, percobaan mengklasifikasikan orientasi keseluruhan dokumen. Orientasi ini dapat berupa positif (default), netral (tidak positif maupun negatif), atau negatif. Perlu dicatat, bagaimanapun bahwa beberapa orang memberikan nilai orientasi ini secara terus menerus atau tidak membedakan. *Fined- grained sentiment analysis* : kategori kedua saat ini sedang mengalami lonjakan popularitas. Implikasinya adalah bahwa akademis memberikan perhatian yang tidak proporsional pada jenis ini. Jika anda ingin mengkategorikan sesuatu, sebaiknya tidak mencarinya ditingkat dokumen, sebaliknya harus mencari bagian tertentu dalam sebuah dokumen.[21]. Pada proses ini menghasilkan berapa banyak data yang diolah termasuk kedalam kategori sentimen analisis komentar positif, negatif dan netral.
4. *Tweet Profiler*, untuk setiap dokumen yang dibagikan oleh tweet, *tweet profiler* mengumpulkan umpan balik pengguna dari server dan menggunakannya untuk menginformasikan rekomendasinya. Alat memberikan informasi ke server pusat, dimana model matematika menghitung kemungkinan berbagai hasil dan intensitas emosi terkait[9]. Intinya fungsi dari tools tweet profiler yang digunakan pada penelitian ini untuk mengukur tingkat emosional yang terkandung dalam isi komentar pada data twitter.

3. HASIL DAN PEMBAHASAN

Proses penelitian ini menggunakan metode *K-Nearest Neighbor* dengan menggunakan program *Orange*. Berikut rangkaian dalam proses penelitian dalam gambar 2. dibawah ini:



Gambar 2. Rangkaian Penelitian.

3.1 Scraping

Proses mendapatkan data opini masyarakat twitter melalui *website* Netlytic.org dengan bantuan kata kunci berupa hastag yaitu #KTTG20 #G20 dan mendapatkan data sebanyak 2500 dari tanggal 15 November 2022 hingga 7 hari kebelakang yang kemudian data tersebut di simpan kedalam format .csv, ditampilkan pada gambar hasil *scraping* data dibawah ini :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	tweetid	guid	link	author	title	description	pubdate	source	favorite	retweet	cl	lang	quoted_tweet
2	1	1.59E+18	https://tw.zulfahmi_ami			Tactical Floor Game telah dijalani	Tactical Floor Game telah dijalankan peranan ti	11/15/2022 1:45	Twitter W	0	0	in		origi
3	2	1.59E+18	https://tw.mishani96			RT @kemendagri: 1. Yang Mulia P.1. Yang Mulia Para Pemimpin, dgn ini sy nyatak		11/15/2022 1:45	Twitter fo	0	2	in		retw
4	3	1.59E+18	https://tw.yeongchawheo			Pelaksanaan KTT G20 Di Bali men	Pelaksanaan KTT G20 Di Bali menjadi puncak da	11/15/2022 1:45	Twitter W	0	0	in		origi
5	4	1.59E+18	https://tw.zain_daud			Polri terus melakukan pengaman	Polri terus melakukan pengamanan penyeleng	11/15/2022 1:45	Twitter W	0	0	in		origi
6	5	1.59E+18	https://tw.poldasulsel			Yang Mulia Para Pemimpin, dgn i	Yang Mulia Para Pemimpin, dgn ini sy nyatakan	11/15/2022 1:45	Twitter fo	0	0	in		origi
7	6	1.59E+18	https://tw.saif1_trg			Polri menggelar latihan secara te	Polri menggelar latihan secara terpusat di Bali	11/15/2022 1:45	Twitter W	0	0	in		origi
8	7	1.59E+18	https://tw.tukangNduleg			RT @Itrenggalek: Perkuat keama	Perkuat keamanan dan Dukung KTTG20 Polri m	11/15/2022 1:45	Twitter fo	0	20	in		retw
9	8	1.59E+18	https://tw.WGunung92			RT @Itrenggalek: Delegasi KTT G	Delegasi KTT G20 telah tiba di Bali, Indonesia. :	11/15/2022 1:45	Twitter fo	0	26	in		retw
10	9	1.59E+18	https://tw.poldasulsel			Yang Mulia Para Pemimpin, dgn i	Yang Mulia Para Pemimpin, dgn ini sy nyatakan	11/15/2022 1:45	Twitter fo	0	0	in		origi
11	10	1.59E+18	https://tw.PemkotaBogor			Presiden RI Joko Widodo pada pi	Presiden RI Joko Widodo pada pidato pembuka	11/15/2022 1:45	Twitter fo	0	0	in		origi
12	11	1.59E+18	https://tw.zain_daud			Polri menggelar latihan secara te	Polri menggelar latihan secara terpusat di Bali	11/15/2022 1:45	Twitter W	0	0	in		origi
13	12	1.59E+18	https://tw.detikinet			Telkomsel menggelar Showcase	Telkomsel menggelar Showcase Pemanfaatan :	11/15/2022 1:45	Echobox	0	0	in		origi
14	13	1.59E+18	https://tw.TanahMerdeka			RT @trivul82: Perdana Menteri	Perdana Menteri Belanda telah tiba di Bali untu	11/15/2022 1:45	Twitter W	0	28	in		retw
15	14	1.59E+18	https://tw.poldasulsel			100 Personel Ditlantas Polda Suls	100 Personel Ditlantas Polda Sulsel ikut amank	11/15/2022 1:45	Twitter fo	0	0	in		origi
16	15	1.59E+18	https://tw.POLSEK_KEERA			Yang Mulia Para Pemimpin, dgn i	Yang Mulia Para Pemimpin, dgn ini sy nyatakan	11/15/2022 1:45	Twitter fo	0	0	in		origi
17	16	1.59E+18	https://tw.ghonik7			RT @Renatha971: Pelabuhan Sai	Pelabuhan Sanur Resmi Dibuka, Presidensi G20	11/15/2022 1:45	Twitter fo	0	4	in		retw
18	17	1.59E+18	https://tw.sylverinus			Terdapat 4 Lokasi penyelenggara	Terdapat 4 Lokasi penyelenggaraan side event	11/15/2022 1:45	Twitter W	0	0	in		origi
19	18	1.59E+18	https://tw.08lutfi			RT @DivHumas_Polri: Yang Mulia	Yang Mulia Para Pemimpin, dgn ini sy nyatakan	11/15/2022 1:45	Twitter fo	0	181	in		retw
20	19	1.59E+18	https://tw.fairiejtt			Sejumlah side event sudah disiaj	Sejumlah side event sudah disiapkan untuk me	11/15/2022 1:45	Twitter W	0	0	in		origi

Gambar 3. Hasil *scraping* data

3.2 Klasifikasi K-Nearest Neighbor

Data yang telah didapatkan dari proses *scraping* kemudian diidentifikasi atau diklasifikasikan ke dalam kategori original, retweet, reply, dan quotes menggunakan metode *K-Nearest Neighbor* dengan bantuan *software* Orange. Data yang telah di klasifikasikan maka mendapatkan hasil popularitas dengan menggunakan *tools* *confusion matrix* seperti pada gambar 4 dibawah :

		Predicted				Σ
		original	quote	reply	retweet	
Actual	original	1107	0	0	0	1107
	quote	0	18	12	3	33
	reply	0	8	49	0	57
	retweet	0	0	0	1303	1303
Σ		1107	26	61	1306	2500

Gambar 4. Confusion Matrix

Hasil dari klasifikasi ternyata hanya 1107 dari 2500 data yang masuk kedalam kategori original yang dimana data tersebut yang akan digunakan pada tahap selanjutnya. Klasifikasi menggunakan metode *K-Nearest Neighbor* mendapatkan tingkat *Accuracy* sebesar 99%, *Precision* sebesar 99%, *F1 Score* sebesar 99%, *Recall* 99% dengan bantuan *tools test and score*.

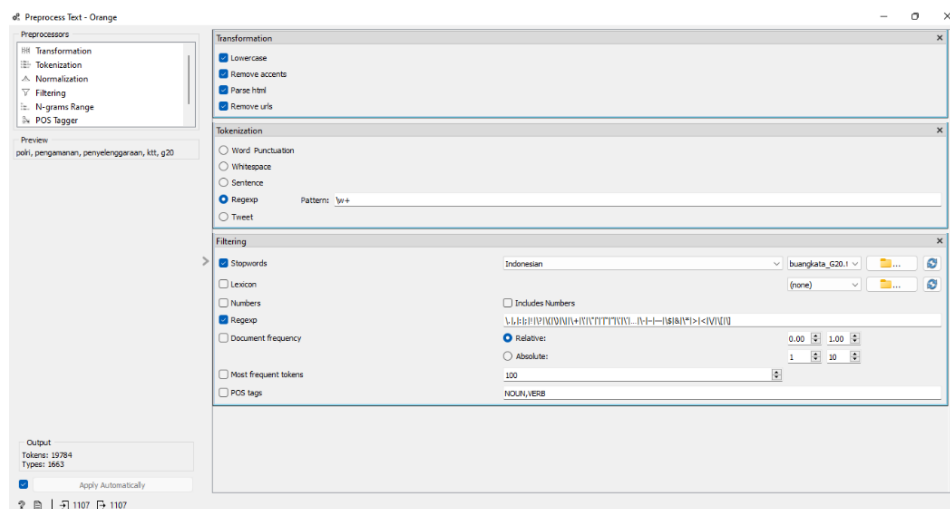
Evaluation results for target (None, show average over classes) ▾

Model	AUC	CA	F1	Precision	Recall
kNN	0.996	0.991	0.990	0.990	0.991

Gambar 5. Test and Score

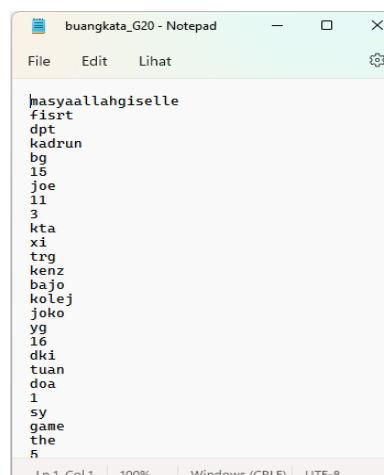
3.3 Pre-processing

Sebelum masuk pada tahap analisis data perlu di persiapkan agar suatu data benar-benar siap dan sesuai dengan yang di harapkan. Langkah yang dilakukan pada tahap *pre-processing* mulai dari *transformation*, *tokenization*, dan *filtering* yang berfungsi sebagai penyaringan atau penghapusan kata, aksen,emoji,link bahkan karakter. Berikut tampilan proses *pre-processing* dibawah ini :



Gambar 6. Pre-processing

Pada fase ini jika masih ada kata yang ingin dihilangkan atau di hapus dapat menambahkan file dengan format .txt dengan penulisan per baris berisi kata yang akan dihilangkan. Contohnya bisa dilihat gambar 7 dibawah :

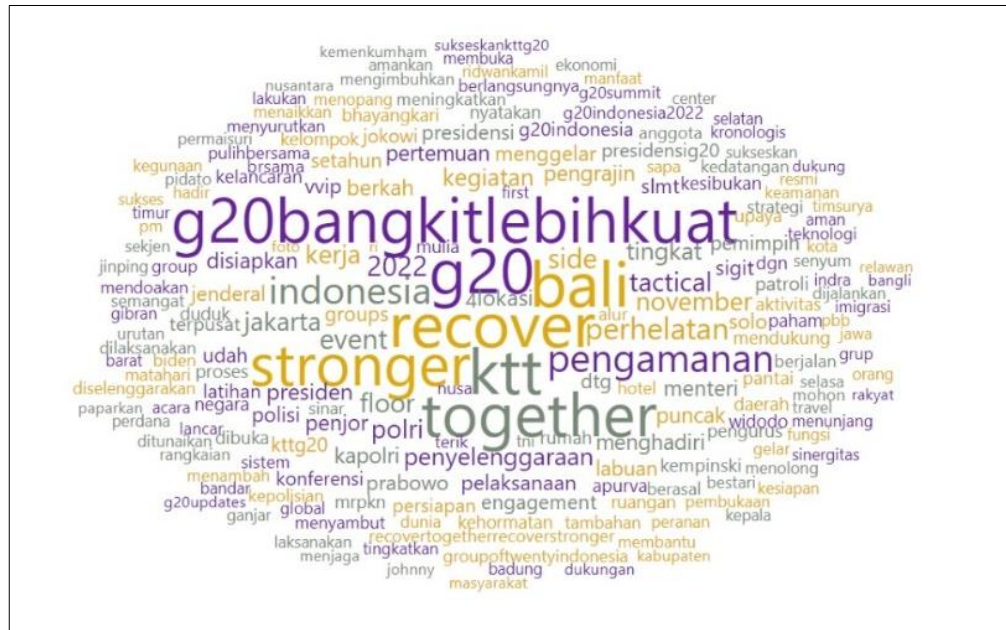


Gambar 7. Contoh Stopword

3.4 Analisis Data

a. Word Cloud

Setelah tahapan *pre-processing* dilakukan, kemudian data di olah menggunakan *word cloud* pada *software orange*. Dengan bantuan *tools word cloud* yang akan menghasilkan pisahan kata yang berbentuk awan, maka dapat diketahui berapa sering muncul kata tersebut digunakan oleh masyarakat dalam beropini mengenai KTT G20 di twitter. Semakin besar kata tersebut bertanda semakin sering muncul kata tersebut didalam data. Hasil *word cloud* sebagai berikut :



Gambar 8. Hasil Word Cloud

Proses dari *word cloud* pada data penelitian menunjukkan beberapa kata yang sering muncul didalam data seperti g20bangkitlebihkuat, g20, bali, recover, stronger, ktt dan stronger.

b. Topic Modeling

Selain pengolahan dengan menggunakan *tools word cloud*, pengolahan juga dilakukan dengan *tools topic modeling* yang berfungsi untuk mengidentifikasi topik-topik yang digunakan masyarakat twitter terhadap KTT G20. Proses ini menghasilkan *topic keyword* yang digunakan masyarakat twitter yang terbagi menjadi 10 topik sebagai berikut :

Topic	Topic keywords
1	recover, g20, ktt, bali, g20bangkitlebihkuat, together, stronger, pengamanan, indonesia, perhelatan
2	recover, g20, indonesia, ktt, bali, stronger, together, g20bangkitlebihkuat, slmt, dtg
3	kerja, pengamanan, indonesia, bali, groups, engagement, pertemuan, menteri, puncak, presidensi
4	slmt, dtg, kerja, 2022, indonesia, november, pengamanan, tingkat, upaya, nyatakan
5	2022, november, bali, terpusat, latihan, persiapan, polri, event, slmt, dtg
6	tactical, floor, 2022, pengamanan, november, kapolri, jenderal, prabowo, sigit, event
7	perhelatan, berkah, penjor, pengrajin, pengamanan, patroli, pantai, kelancaran, daerah, bhayangkari
8	patroli, pantai, perhelatan, presiden, kapolri, menghadiri, jenderal, prabowo, sigit, berkah
9	presiden, daerah, pengurus, bhayangkari, kelancaran, kegiatan, kttg20, kempinski, apurva, berkah
10	disiapkan, penyelenggaraan, jakarta, event, side, 4, solo, labuan, kegiatan, lokasi

Gambar 9. Topic Modeling

c. Sentiment Analysis

Tahapan analisis menggunakan algoritma vader yang telah disediakan pada *tools software orange*, bertujuan untuk mengklasifikasikan data yang di olah masuk kedalam opini negatif, positif atau netral. Vader memberi peringkat dan menetapkan titik teks berdasarkan nilai setiap kata yang tercantum dalam kamus Vader. Hasil akhir penilaian adalah skor total yang dijumlahkan. Hasil analisis sentimen akan dihitung bobot positif, negatif, netral dengan melihat dari total skor (*compound*) dengan format data numerik.

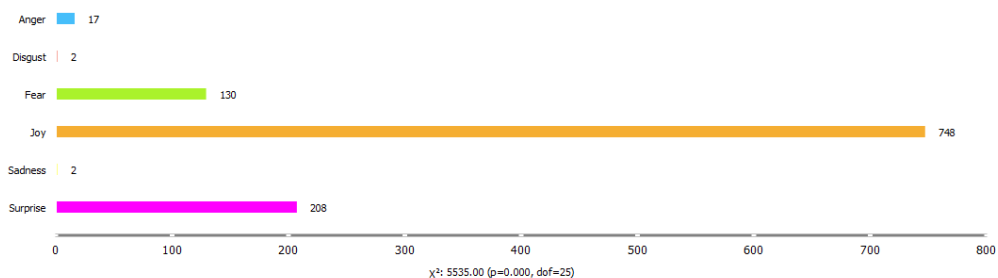
	source	description	author	pos	neg	neu	compound
64	Twitter Web App	Pelaksanaan K.L...	mfaizul_h	0.157	0	0.843	0.6808
65	Twitter Web App	Pelaksanaan K.L...	hasan_sidek	0.157	0	0.843	0.6808
66	Twitter Web App	Pelaksanaan K.L...	marziahrosli	0.157	0	0.843	0.6808
67	Twitter Web App	Pelaksanaan K.L...	jamalsahir90	0.157	0	0.843	0.6808
68	Twitter Web App	Pelaksanaan K.L...	poqzqqlrv	0.157	0	0.843	0.6808
69	Twitter Web App	Pelaksanaan K.L...	hini_roshini	0.157	0	0.843	0.6808
70	Twitter Web App	Pelaksanaan K.L...	syaidjtt	0.157	0	0.843	0.6808
71	Twitter Web App	Pelaksanaan K.L...	yahyajt	0.157	0	0.843	0.6808
72	Twitter Web App	Pelaksanaan K.L...	oldapaza	0.157	0	0.843	0.6808
73	Twitter Web App	Pelaksanaan K.L...	zfadhli31	0.157	0	0.843	0.6808
74	Twitter Web App	Pelaksanaan K.L...	samsuljtt	0.157	0	0.843	0.6808
75	Twitter Web App	Pelaksanaan K.L...	ZSiampapaga	0.157	0	0.843	0.6808
76	Twitter Web App	Pelaksanaan K.L...	nik4_trg	0.157	0	0.843	0.6808
77	Twitter Web App	Pelaksanaan K.L...	rahman5_trg	0.153	0	0.847	0.6808
78	Twitter Web App	Pelaksanaan K.L...	kamal_sheela	0.153	0	0.847	0.6808
79	Twitter Web App	Pelaksanaan K.L...	sharifah8_trg	0.153	0	0.847	0.6808
80	Twitter Web App	Pelaksanaan K.L...	HazrinSyafiq	0.153	0	0.847	0.6808
81	Twitter Web App	Pelaksanaan K.L...	nasir14_trg	0.153	0	0.847	0.6808
82	Twitter Web App	Ini Keistimewaa...	mas_iko	0.259	0	0.741	0.6369
83	Twitter Web App	Ini Keistimewaa...	herman_joseph	0.219	0	0.781	0.6369
84	Twitter Web App	Polri mengela...	RESKUNINGAN	0.057	0.048	0.895	0.1027
85	Twitter for Andr...	Yang Mulia Par...	POLSEK_KEERA	0	0	1	0
86	Twitter Web App	HERE WE GO !!! ...	notife21	0	0	1	0
87	Twitter Web App	Keknya hampir ...	BNatawijaya	0	0	1	0
88	Twitter Web App	Para menteri Ka...	kompascom	0	0	1	0
89	Twitter Web App	HERE WE GO !!! ...	Keong_Dewaku	0	0	1	0
90	Twitter for Andr...	Quotes Preside...	b9352uyy	0	0	1	0
91	Twitter for Andr...	Senyum dan Sa...	Bhabink23994533	0	0	1	0
92	Twitter for Andr...	Senyum dan Sa...	KasimanHumas	0	0	1	0
93	Twitter for Andr...	Polisi Perairan S...	FirmanRiza6	0	0	1	0
94	Twitter for Andr...	Polisi Perairan S...	FirmanRiza6	0	0	1	0

Gambar 10. Hasil analisis sentimen

Berdasarkan hasil dari analisis sentimen menggunakan algoritma vader terdapat opini masyarakat twitter yang diklasifikasikan positif sebanyak 404 data, negatif sebanyak 89 data, dan netral sebanyak 614 data.

d. Tweet Profiler

Tweet profiler digunakan untuk menghitung tingkat emosional masyarakat yang beropini terhadap KTT G20. Terdapat 6 kategori emosional berupa *anger* (amarah), *disgust* (menjijikan), *fear* (takut), *joy* (sukacita/gembira), *sadness* (sedih), *surprise* (kejutan). Pada *tools* ini menyediakan 3 type yaitu *Ekmans*, *Plutchiks*, *Profile of Mood States* (POMS). Pada penelitian ini menggunakan kategori *Ekmans* (*Multi Class*) yang akan di implementasikan pada 1107 tweet yang beropini mengenai KTT G20. Dengan dihubungkan pada *tools boxplot* maka hasil dari analisis tingkat emosional ini terlihat sebagai berikut :



Gambar 11. Bloxpot tingkat emosional

Hasil dari perhitungan 6 kategori emosional terhadap opini masyarakat mengenai KTT G20 menggunakan *tools tweet profiler* yang kemudian dihubungkan dengan *tools bloxpot* menunjukkan bahwa kebanyakan masyarakat beropini dengan tingkat emosional *joy*, *surprise*, dan *fear*.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan dan penjelasan yang diberikan, kita dapat menarik kesimpulan tentang sentimen populasi twitter terhadap opini terhadap KTT G20 dengan menggunakan data yang diambil dari 2500 data menggunakan teknik *scraping*. Dengan metode *K-nearest neighbor* maka dapat diketahui 2500 data tweet masuk kedalam klasifikasi original, reply, quotes, retweet dengan tingkat *Accuracy* sebesar 99%, *Precision* sebesar 99%, *F1 Score* sebesar 99%, *Recall* sebesar 99% dengan hasil 1107 data yang masuk ke dalam klasifikasi tweet original yang kemudian data tersebut akan di olah ke dalam tahap selanjutnya. *Tools word cloud* dan *topic modeling* digunakan untuk menentukan kata dan topik mana yang paling sering ditemui pada data yang sedang di uji. Analisis dengan menggunakan *tools analisis sentimen* dan *tweet profiler* yang dilakukan terhadap 1107 tweet original mendapatkan hasil 89 tweet Negatif, 614 Neutral, 404

positif dengan mayoritas tingkat emosional *Joy*, *Surprise* dan *Fear*. Hasil penelitian tersebut di dapatkan dengan cara mengolah data pada *software orange*.

REFERENCES

- [1] W. R. D. Astuti and L. M. Fathun, "Diplomasi Ekonomi Indonesia di dalam Rezim Ekonomi G20 pada Masa Pemerintahan Joko Widodo," *Intermestic J. Int. Stud.*, vol. 5, no. 1, p. 47, 2020, doi: 10.24198/intermestic.v5n1.4.
- [2] H. R. Dhega *et al.*, "BALI," vol. 3, no. 2, pp. 100–115.
- [3] I. G. J. Pratama, W. Ardani, and I. A. S. Putri, "Pemanfaatan Presidensi G20 sebagai Sarana Marketing dan Branding Ekonomi Kreatif pada Era Pandemi Covid-19," *Lensa Ilm. J. Manaj. dan Sumberd.*, vol. 1, no. 1, pp. 28–33, 2022, doi: 10.54371/jms.v1i1.164.
- [4] P. N. P. Santi, W. Ardani, and I. A. S. Putri, "Presidensi G20 sebagai Sarana Marketing dan Branding Pariwisata Indonesia serta Pengaruhnya terhadap Peningkatan Kunjungan Wisatawan pada Era Pandemi Covid-19 (Studi Kasus di Hotel Melia Bali)," *Lensa Ilm. J. Manaj. dan Sumberd.*, vol. 1, no. 1, pp. 15–20, 2022, doi: 10.54371/jms.v1i1.167.
- [5] J. Indonesia, "AKTUALISASI INDONESIA DALAM G20 : PELUANG ATAU TREN ? Sukma Sushanti Program Studi Hubungan Internasional Fakultas Ilmu Sosial dan Ilmu Politik , Universitas Udayana Email : sukmasushanti@gmail.com PENDAHULUAN Implikasi dari terbentuknya G20 semakin kuat ,” pp. 1–14, 2022.
- [6] M. Kartika, S. Saepudin, and D. Gustian, "Analisis Sentimen Dampak Covid-19 Terhadap Pembatalan Keberangkatan Ibadah Haji Pada Tahun 2020," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 964–972, 2021.
- [7] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, p. 32, 2021, doi: 10.24014/coreit.v7i1.12945.
- [8] R. Amelia, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *J. Ris. Komput.*, vol. 9, no. 2, pp. 338–343, 2022, doi: 10.30865/jurikom.v9i2.3895.
- [9] S. Indah Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald’Sxbts Menggunakan Orange Data Mining," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 28–35, 2021.
- [10] S. S. Hilabi, H. Prabowo, H. L. H. S. Warnars, and T. N. Mursitama, "Recommendation System Model for Political Communication through Social Media using Content Based Filtering Method," *Int. J. Appl. Eng. Technol.*, vol. 4, no. 1, pp. 34–39, 2022.
- [11] M. F. El Firdaus, N. Nurfaizah, and ..., "Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *JURIKOM (Jurnal ...)*, vol. 9, no. 5, pp. 1329–1336, 2022, doi: 10.30865/jurikom.v9i5.4774.
- [12] L. R. Dharmawan, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen pada Sosial Media Twitter Terhadap Layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya dengan Metode K- Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 959–965, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7099>.
- [13] R. Rasenda, H. Lubis, and R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 369, 2020, doi: 10.30865/mib.v4i2.2051.
- [14] F. Rizqi Irawan, "Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearset Neighbors," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.33387/jiko.v5i1.4267.
- [15] A. Asro'i and H. Febuariyanti, "Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan PpkM Menggunakan Metode K-Nearest Neighbor," *J. Khatulistiwa Inform.*, vol. 10, no. 1, pp. 17–24, 2022, doi: 10.31294/jki.v10i1.12624.
- [16] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, "K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 2, pp. 152–158, 2022, doi: 10.37859/coscitech.v3i2.3841.
- [17] S. Analisis Kegiatan Trading dengan SVM, K. R. Dan, N. Resti Wardani, S. Saepudin, and C. Warman, "Sentimen Analisis Kegiatan Trading Pada Ap-likasi Twitter dengan Algoritma SVM, KNN Dan Random Forrest," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 2, pp. 863–870, 2022.
- [18] P. Arsi, I. Prayoga, and M. H. Asyari, "Klasifikasi Sentimen Publik Terhadap Jenis Vaksin Covid-19 yang Tersertifikasi WHO Berbasis NLP dan KNN," vol. 7, pp. 260–266, 2023, doi: 10.30865/mib.v7i1.5418.
- [19] R. Saptari, Rianto, and A. I. Gufroni, "Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Unit Gawat Darurat Rumah Sakit Umum di Indonesia Menggunakan Seleksi Fitur Information Gain dan Support Vector Machine," *J. Oof Informatics Educ.*, vol. XX, pp. 104–110, 2018, [Online]. Available: <http://e-journal.ivet.ac.id/index.php/jiptika/article/view/1925/1369>.
- [20] J. A. Josen Limbong, I. Sembiring, K. Dwi Hartomo, U. Kristen Satya Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor Meth," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2019, doi: 10.25126/jtiik.202294960.
- [21] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, "Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.