

Analisis Sentimen Pengguna Twitter terhadap Konflik Rusia-Ukraina Menggunakan Naïve Bayes dan Lexicon Based Features

Barnes J. Manurung¹, Bitu Parga Zen^{2*}, Yohani Setiya Rafika Nur³, Roland Claudio Felle⁴, Eryan Ahmad Firdaus⁵

^{1,3}Program Studi Teknik Informatika, Universitas Telkom, Indonesia

²Program Studi Teknik Informatika, Universitas Ma Chung, Indonesia

^{4,5}Program Studi Informatika, Universitas Pertahanan Republik Indonesia, Indonesia

Email: ¹barnes.manurung@gmail.com, ²bitu.parga@machung.ac.id, ³yohani@telkomuniversity.ac.id, ⁴rolandfelle14@gmail.com, ⁵eryan.firdaus@idu.ac.id

ABSTRAK – Konflik antara Rusia dan Ukraina merupakan salah satu isu internasional yang terus menjadi perhatian dunia. Konflik ini bermula sejak 2014, saat pelengseran Presiden Yanukovych menyebabkan perpecahan antara pendukung Uni Eropa dan pendukung Rusia di Ukraina. Ketegangan meningkat tajam pada akhir 2021 hingga akhirnya Presiden Rusia, Vladimir Putin, memutuskan melakukan agresi militer ke Ukraina pada 24 Februari 2022. Twitter sebagai media sosial menjadi wadah utama bagi masyarakat untuk menyuarakan opini mereka terhadap konflik ini. Penelitian ini bertujuan untuk menganalisis sentimen publik di Twitter mengenai konflik Rusia-Ukraina. Metode yang digunakan adalah kombinasi algoritma Naïve Bayes dan pendekatan Lexicon Based Features. Naïve Bayes digunakan untuk mengklasifikasikan data sentimen, sedangkan Lexicon Based Features berfungsi untuk menentukan bobot sentimen positif maupun negatif dalam teks. Hasil penelitian menunjukkan bahwa metode kombinasi ini efektif dalam mengkategorikan opini publik berdasarkan sentimen mereka terhadap konflik tersebut. Analisis ini dapat memberikan gambaran umum mengenai persepsi global terhadap konflik Rusia-Ukraina dan menjadi bahan pertimbangan dalam memahami opini publik dunia maya.

Kata Kunci: Konflik Rusia-Ukraina, Analisis Sentimen, Naïve Bayes, Lexicon Based Features, Twitter.

ABSTRACT – The conflict between Russia and Ukraine remains one of the major international issues drawing global attention. It began in 2014 with the ousting of President Yanukovych, which triggered a political divide between pro-European Union and pro-Russian factions within Ukraine. Tensions escalated significantly by late 2021, culminating in Russia's military aggression against Ukraine on February 24, 2022, under President Vladimir Putin's directive. Twitter, as a social media platform, became a major outlet for the public to express their opinions regarding the conflict. This study aims to analyze public sentiment on Twitter concerning the Russia-Ukraine conflict. The methods used in this research are a combination of the Naïve Bayes algorithm and Lexicon Based Features. Naïve Bayes is utilized to classify sentiment data, while Lexicon Based Features assign weights to positive and negative sentiment words in textual data. The results show that this combined method is effective in categorizing public opinions based on their sentiments towards the conflict. This sentiment analysis provides a broader understanding of global perceptions and may serve as a reference for assessing public opinion in digital media contexts.

Keywords: Russia-Ukraine Conflict, Sentiment Analysis, Naïve Bayes, Lexicon-Based Features, Twitter.

PENDAHULUAN

Perkembangan teknologi saat ini semakin cepat, khususnya smartphone yang berbasis android. Adanya smartphone yang berbasis android ini berdampak juga semakin banyaknya media sosial yang bermunculan. Salah satu media umum yang masih sering di gunakan masyarakat merupakan Twitter. Peringkat media umum sesuai jumlah pengguna pada Indonesia Twitter berada di

posisi ke lima sesudah YouTube, WhatsApp, Facebook serta Instagram [1]. Twitter merupakan salah satu media sosial yang menjadi wadah bagi publik buat mengutarakan opini mereka terhadap topik yang dibicarakan. sebagai wadah opini publik, terkumpul berbagai data tweet pengguna. Data ini bisa diolah serta dianalisis buat mengetahui sentimen pengguna Twitter terhadap peristiwa atau topik yang sedang terjadi pada saat

ini [2]. Topik yang sedang menyita perhatian pengguna twitter adalah konflik Rusia dan Ukraina. Konflik Russia dan Ukraina menjadi salah satu perseteruan yang menyita perhatian dunia internasional hingga kini. Pada Desember 2021 Ketegangan semakin meningkat, dengan lebih dari 100.000 tentara Rusia ditempatkan di perbatasan, beserta dengan artileri berat.

Pembicaraan antara Rusia dan sekutu barat sejauh ini tidak berhasil, dengan Presiden Putin teguh dalam tuntutananya supaya Ukraina tidak bergabung dengan NATO, dan Barat tidak mau mengakuinya [3]. Opini masyarakat Indonesia semakin memuncak dengan adanya perseteruan konflik tersebut. Opini publik tersebut memiliki berbagai pandangan baik positif, netral, maupun negatif terhadap ke 2 belah pihak. Pandangan masyarakat tadi dituangkan di berbagai media sosial. Salah satu media sosial yang banyak digunakan ialah Twitter [4]. Data ini dapat diolah dan dianalisis untuk memahami sentimen pengguna Twitter terhadap konflik tersebut. Penelitian mengenai analisis sentimen terhadap data tweet pengguna Twitter telah banyak dilakukan oleh berbagai peneliti terdahulu. Penelitian yang dilakukan tersebut banyak menggunakan metode Naïve Bayes dan Lexicon Based Features [5]. Penelitian yang dilakukan oleh Ali Sholihin, Haviluddin, Novianti Puspitasari, Masna Wati, dan Islamiyah yang, yang menghasilkan iterasi tertinggi 75% [6]. Ada juga penelitian dari Rifiana Arief dan Karel Imanuel yang membahas sentimen topik viral desa penari pada media sosial twitter dengan menggunakan metode Lexicon Based dan mendapatkan hasil analisa dari 1000 tweet 33 tweet positif (3.3%), 767 tweet netral (76.7%) dan 200 tweet negatif (20%) [7].

Berdasarkan penelitian di atas, penulis melakukan pengolahan mengenai analisis sentimen dari data tweet pengguna terhadap konflik Rusia dan Ukraina. Penelitian ini menggunakan Analisis Sentimen artinya proses mengekstraksi, mengolah data maupun memahami suatu data tekstual secara otomatis [8]. Peneliti menggunakan metode Naïve Bayes dan Lexicon Based Features yang digunakan buat menganalisis sentimen dari data tweet terhadap konflik Rusia dan Ukraina. Naïve Bayes merupakan salah satu algoritma machine learning yang sering diimplementasikan dalam analisis sentimen. Sedangkan Lexicon Based Features menentukan bobot setiap kata sentimen positif atau sentimen negatif, orientasi sentimen suatu kata dalam yang sering diimplementasikan dalam

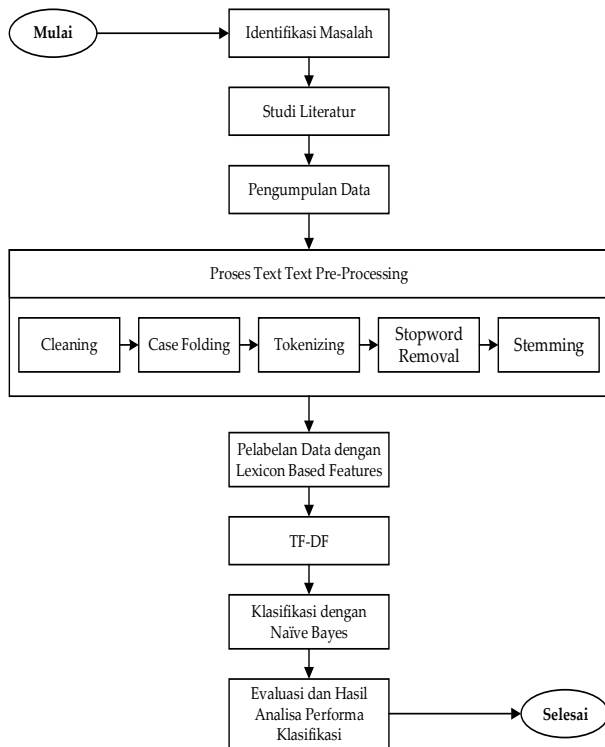
analisis sentimen. Pada penelitian ini peneliti menggunakan proses text mining, yaitu suatu teknik yang dipergunakan buat menangani persoalan klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi dan mempunyai sifat hampir sama dengan data mining namun hanya penekanan pada teks daripada bentuk data yang lebih terstruktur [9]. Penggunaan teknologi *Natural Language Processing* (NLP), dengan menerapkan prinsip-prinsip linguistik computational dalam menguraikan serta menginterpretasikan set data.

Natural Language Processing berkaitan dengan personal komputer bisa dipergunakan untuk memahami serta memanipulasi teks bahasa alami (natural language) untuk mendapatkan informasi eksklusif [10]. Dalam penelitian ini peneliti juga menggunakan TF-IDF (Term Frequency-Invers Document Frequency) menjadi seleksi fitur [11]. Pada penelitian yang dilakukan oleh Bitu Parga Zen, Irwan Susanto, dan Dian Finaliamartha pada tahun 2021 yang berjudul "TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News" [12]. Berdasarkan penjelasan di atas, penelitian ini akan dituangkan dalam Laporan Akhir yang berjudul "Analisis Sentimen Pengguna Twitter terhadap Konflik Rusia-Ukraina Menggunakan Naïve Bayes dan Fitur Lexicon". Pemilihan judul ini didasarkan pada tujuan untuk menganalisis tanggapan pengguna Twitter terhadap konflik Rusia dan Ukraina secara cepat dan akurat, menggunakan dua metode tersebut. Penelitian ini diharapkan dapat memberikan kontribusi baru, mengingat belum ada penelitian yang menggunakan kombinasi metode tersebut pada data Twitter terkait konflik ini.

METODE

Dalam melakukan analisis sentimen, terdapat beberapa tahapan penelitian yang perlu dilalui untuk memastikan hasil pengujian yang akurat dan valid. Setiap tahapan memiliki peran penting dalam mendukung proses analisis, mulai dari pengumpulan data, pemrosesan data, hingga evaluasi hasil. Tahapan-tahapan ini dirancang secara sistematis untuk memberikan gambaran yang jelas dan menyeluruh mengenai alur penelitian, serta untuk memudahkan pelaksanaan analisis secara efisien. Dengan mengikuti alur yang telah ditetapkan, peneliti dapat memastikan bahwa setiap langkah dilakukan dengan tepat. Tahapan penelitian yang digunakan dalam studi ini dapat

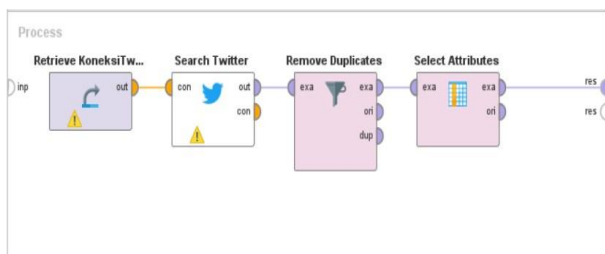
dilihat pada gambar 1 di bawah ini, yang menggambarkan urutan tahapan serta hubungan antar tahapan yang saling berkaitan.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang dikumpulkan berupa tweet pada Twitter tentang konflik Rusia dan Ukraina. Mengumpulkan data ini bertujuan agar mendapatkan informasi tentang konflik Rusia dan Ukraina ini untuk dapat di analisis dengan cara klasifikasi menggunakan metode Naïve Bayes dan Lexicon Based Features. Dalam mengumpulkan data, peneliti melakukan dengan cara crawling menggunakan Rapid Miner.



Gambar 2. Operator Rapid Miner

2.2. Text Pre-Processing

Setelah melakukan pengumpulan data tweet pada data maka selanjutnya langkah untuk text pre-processing.

1. Cleaning

Dengan cara kata atau karakter yang akan dihilangkan seperti simbol, link URL, tagar (#),

nama pengguna atau mention (@namaakun), emoticon dan RT (Retweet). Contoh proses Cleaning dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Cleaning

Sebelum melakukan Cleaning	Setelah melakukan Cleaning
Invasi #Rusia ke #Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil di kepung nya.	Invasi Rusia ke Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil di kepung nya.

2. Case Folding

Case Folding adalah mengubah semua karakter huruf menjadi huruf kecil (lowercase). Contoh proses Case Folding dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Case Folding

Sebelum melakukan Case Folding	Setelah melakukan Case Folding
Invasi #Rusia ke #Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil dikepungnya.	invasi rusia ke ukraina telah memasuki hari ke-12. kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara rusia di kota-kota ukraina yang telah berhasil di kepung nya.

3. Tokenizing

Tokenizing yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata.

Tabel 3. Tokenizing

Sebelum melakukan Tokenizing	Setelah melakukan Tokenizing
Invasi Rusia ke Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang	['Invasi', 'Rusia', 'ke', 'Ukraina', 'telah', 'memasuki', 'hari', 'ke-12', 'Kyiv', 'mengatakan', 'ada', 'eskalasi', 'penembakan', 'yang', 'dilakukan', 'tentara',

dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil dikepungnya.	'Rusia', 'di', 'kota-kota', 'Ukraina', 'yang', 'telah', 'berhasil', 'dikepungnya']
---	--

4. Stopword Removal

Stopword Removal yaitu proses penghapusan kata-kata yang terdapat pada stoplist. Stoplist itu sendiri berisi kosakata-kosakata yang bukan merupakan ciri dari suatu dokumen dan bertujuan agar menghilangkan kata yang tidak berguna.

Tabel 4. Stopword Removal

Sebelum melakukan Stopword Removal	Setelah melakukan Stopword Removal
Invasi Rusia ke Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil dikepungnya.	Invasi Rusia Ukraina memasuki hari ke-12. Kyiv mengatakan eskalasi penembakan dilakukan tentara Rusia kota-kota Ukraina telah berhasil dikepungnya.

5. Stemming

Stemming ialah suatu proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (stem). Stemming bertujuan untuk menghilangkan berbagai imbuhan seperti prefiks, sufiks, maupun konfiks yang ada pada setiap kata.

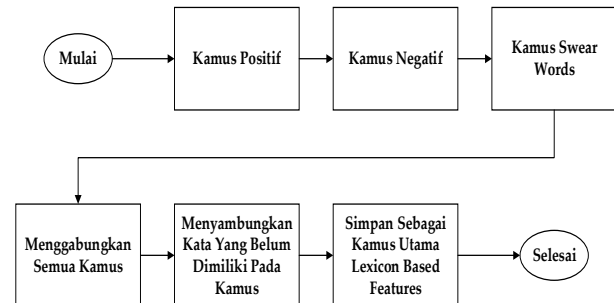
Tabel 5. Stemming

Sebelum melakukan Stemming	Setelah melakukan Stemming
Invasi Rusia ke Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil dikepungnya.	invasi rusia ke ukraina telah masuk hari ke-12 kyiv kata ada eskalasi tembak yang laku tentara rusia di kota ukraina yang telah hasil kepung

2.3. Pelabelan Data

Setelah melakukan pre-processing data, data tersebut masih bersifat mentah maka dibutuhkan label pada data tersebut dan bertujuan agar bisa melakukan klasifikasi. Pada proses pelabelan menggunakan lexicon, akan dibagi menjadi 2 tahap yaitu proses pembuatan kamus lexicon dan

pelabelan dataset. lexicon based sendiri adalah sebuah metode pelabelan yang menggunakan kamus utama lexicon sebagai acuan untuk menghitung bobot atau nilai polaritas dari tiap kata opini atau sentimen. Berikut adalah alir pembuatan kamus Lexicon Based Features.



Gambar 3. Tahap Pelabelan Data

Kamus utama ini nantinya akan dibentuk dari kamus-kamus kecil lexicon, ada juga kamus didapat dari repository GitHub. Untuk kamus kecil yang terdiri dari kamus lexicon positif dan lexicon negatif berasal dari repository GitHub Inset (Indonesia Sentiment Lexicon) milik Fajri Koto, and Gemala Y. Rahmaningtyas.

2.4. Pembobotan TF-IDF

Setelah tahapan pelabelan data berhasil dilakukan pada data, maka tahapan selanjutnya adalah ekstraksi fitur menggunakan skema Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF merupakan proses pembobotan yang dilakukan dengan cara menghitung frekuensi kemunculan suatu kata pada tiap dokumen (TF) dan frekuensi kemunculan suatu kata pada keseluruhan dokumen (IDF). Pembobotan TF-IDF bertujuan untuk mereduksi ukuran panjang dokumen yang awalnya beragam sehingga berubah ke ukuran yang tetap.

2.5. Klasifikasi Menggunakan Metode Naive Bayes dan Lexicon Based Features

Untuk melakukan klasifikasi sentimen akan menggunakan data hasil proses pre-processing, sampai dengan pembobotan kata dengan TF-IDF. Setelah data berhasil di-training kemudian akan dilakukan pengujian menggunakan data test untuk menguji hasil ketepatan klasifikasi yang dilakukan.

2.6. Evaluasi dan Analisa Performa Klasifikasi

Pada langkah evaluasi ini parameter yang digunakan adalah confusion matrix. Tujuan dari adanya evaluasi ini agar dapat melihat nilai accuracy, precision, recall dan f1 score dengan melihat performa model klasifikasi Naive Bayes

untuk kasus konflik Rusia dan Ukraina pada Twitter. Setelah mendapatkan hasil dari evaluasi maka dapat di tarik kesimpulan dari penelitian ini.

HASIL DAN PEMBAHASAN

Pengumpulan data dalam penelitian dilakukan dengan crawling yang menggunakan tools RapidMiner dengan beberapa operator yang terhubung dengan API Twitter. Berdasarkan hasil pengumpulan data 2003 data twitter dapat di lihat pada tabel 6 di bawah.

Tabel 6. Hasil Pengumpulan Data

Text
Heboh Konflik Ukraina-Rusia, Gigi Hadid Ingatkan Masih Ada Perang di Palestina #TempoCantika
"Invasi #Rusia ke #Ukraina telah memasuki hari ke-12. Kyiv mengatakan ada eskalasi penembakan yang dilakukan tentara Rusia di kota-kota Ukraina yang telah berhasil dikepungnya. https://t.co/ZeYBDxJjnO "
Keluarga dari negara konflik di Suriah dan Palestina yang memulai hidup baru di Ukraina membicarakan tentang keterkejutan mereka saat terperangkap di perang yang lainnya. https://t.co/wjRNIXrfG2
RT @skidapapaw: @TxdariHI Baca dmn ya aku lupa. Tentara ukraina yg ketangkep di medan perang oleh tentara rusia lgs dikasih mkn+minum...

Data yang diperoleh, akan melalui proses pre-processing. Proses ini digunakan untuk mengubah data yang masih mentah dan tidak terstruktur menjadi data yang siap untuk diolah. Dataset ini juga telah melewati proses menghilangkan duplikat data. Duplikat adalah proses penghapusan duplikat kata ini yang semula dataset berjumlah 2003 data tweet menjadi 1860 data tweet. Dataset ini sudah siap digunakan untuk proses penelitian. Proses yang dilakukan setelah pre-processing adalah proses pelabelan atau pemberian label kelas tiap data tweet.

Tabel 7. Hasil Pre-Processing

Tweet
heboh konflik ukraina rusia gigi hadid ingatkan perang palestina
invasi memasuki kyiv eskalasi penembakan tentara rusia kota kota ukraina berhasil dikepungnya
keluarga negara konflik suriah palestina memulai ukraina membicarakan keterkejutan terperangkap perang

baca dmn lupa tentara ukraina ketangkep medan perang tentara rusia lgs dikasih mkn minum

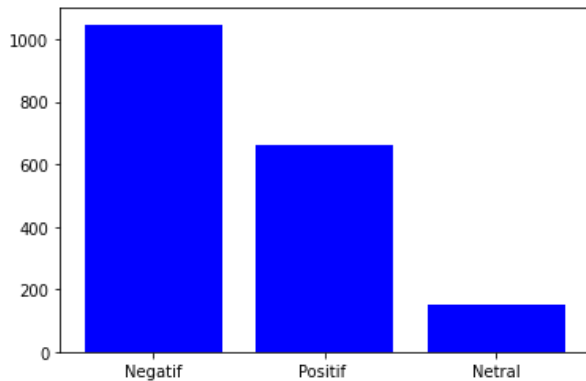
Setelah itu adalah proses pelabelan dataset dilakukan setelah proses pre-processing menggunakan metode Lexicon Based features dengan menggunakan kamus lexicon yang diperoleh dari repository github.com milik evanmertua34 "<https://github.com/evanmertua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based/tree/master/lexicon>". Dalam repository miliknya, terdapat berbagai macam kamus lexicon yang terdiri dari kamus inset negatif, kamus inset positif, kamus-kamus tersebut juga di tambahkan beberapa kata yang tidak dimiliki sebelumnya oleh peneliti agar pelabelan bisa berjalan dengan baik, dan beberapa kamus lexicon lain yang nantinya akan disatukan terlebih dahulu untuk membuat sebuah kamus utama lexicon yang memiliki total kata berjumlah +11110 dengan variasi bobot kata mulai dari -5 sampai 5. Dataset yang berjumlah 1860 tweet hasil dari pre-processing diberikan label sesuai bobot kata yang ada di setiap tweet.

Tabel 8. Hasil Pelabelan Data Tweet

Tweet	Bobot Tweet
heboh konflik ukraina rusia gigi hadid ingatkan perang palestina	-4
invasi memasuki kyiv eskalasi penembakan tentara rusia kota kota ukraina berhasil dikepungnya	-4
keluarga negara konflik suriah palestina memulai ukraina membicarakan keterkejutan terperangkap perang	-1
baca dmn lupa tentara ukraina ketangkep medan perang tentara rusia lgs dikasih mkn minum	3
heboh konflik ukraina rusia super model gigi hadid ingatkan perang palestina,	-3
ihsg diprediksi hijau dibayangi perang rusia ukraina	5
serukan umat muslim tak ikutan isis nyatakan perang rusia ukraina azab ilahi hancurkan dunia barat	5
jokowi desak rusia ukraina gencatan senjata setop perang	-7

Berdasarkan penjelasan di atas, hasil akumulasi polaritas kata di tiap data yang bernilai > 0, maka data tweet tersebut menandakan positif

dan akan di beri label positif, Namun jika data tersebut bernilai < 0 , maka data tersebut diberi label negatif. Pada tahap pelabelan, akan otomatis ini didapatkan data sebanyak 1047 tweet bernilai negatif, 662 tweet bernilai positif, dan 151 tweet bernilai netral. Hasil akumulasi data bernilai positif, negatif, dan netral dapat dilihat pada Gambar di bawah ini:



Gambar 4. Diagram Hasil Label Lexicon Based

Setelah melakukan pelabelan dengan lexicon based features, dataset kemudian dibagi menjadi dua bagian yaitu data train dan data test dengan rasio atau komposisi perbandingan 80% untuk data train dan 20% untuk data test. Pada tabel 9 di bawah menunjukkan distribusi dari hasil pembagian data train dan data test.

Tabel 9. Pembagian Data Train dan Data Test

Label	Pembagian Dataset (80%:20%)		
	Data Train	Data Test	Total
Positif	525	137	662
Negatif	837	210	1047
Netral	126	25	151
Total Label Data			1860

Selanjutnya, data train dan data test memasuki proses pembobotan kata dengan menggunakan TF-IDF. TF-IDF merupakan proses pembobotan kata pada dokumen yang didasari oleh frekuensi kemunculan kata terhadap dokumen tersebut. Proses TF-IDF berfungsi untuk mendapatkan nilai vector, nilai vector inilah yang akan di proses ke dalam vector space, yang kemudian di proses menggunakan Naive Bayes Classifier. Vector space yang dibangun dari data train memiliki ukuran (1488), dimana 1488 adalah jumlah dokumen atau tweet atau yang sering disebut sebagai corpus yang masuk ke dalam proses TF-IDF data train. Untuk data test, memiliki ukuran (372).

Tabel 10. Data Train

Tweet	Term	Bobot
0	4689	0.1378778602382717
0	4670	0.1826436426618855
0	3756	0.1749473725098706
0	3717	0.1495686841827171
0	3243	0.0446518986704978
:	:	:
1487	2585	0.3853792743669896
1487	1766	0.2607101828958915
1487	1364	0.4510451611396212
1487	1078	0.5245965462896716
1487	386	0.5245965462896716

Tabel 10 memberikan hasil dari TF-IDF data train yang telah dilakukan. Kolom tweet berisi baris sebanyak jumlah data di data train. Dengan kata lain, menunjukkan index dari tiap dokumen. Lalu kolom term berisi index dari feature word yang dihasilkan dari bag of words. Pada kolom bobot berisi bobot dari tiap term yang dihasilkan. Cara membaca tabel 10 adalah index dokumen ke-0 (menggunakan konsep array) mengandung term ke 4689 dengan bobot kata 0.1378778602382717.

Tabel 11. Data Test

Tweet	Term	Bobot
0	4706	0.0859668477174502
0	4677	0.4116861982416009
0	4557	0.3539654786867529
0	4532	0.1935919599550111
0	3537	0.1215905016429799
:	:	:
371	4299	0.5015744784975856
371	3717	0.0860471827643372
371	3243	0.1027533298609833
371	2677	0.2274770153462850
371	1962	0.4079936139030151

Tabel 11 memberikan hasil dari TF-IDF data test yang telah dilakukan. Kolom tweet berisi baris sebanyak jumlah data di data train. Dengan kata lain, menunjukkan index dari tiap dokumen. Lalu kolom term berisi index dari feature word yang dihasilkan dari bag of words. Pada kolom bobot berisi bobot dari tiap term yang dihasilkan. Cara membaca tabel 11 adalah dokumen dengan index ke 0 yang mengandung term ke 4706 dengan bobot term 0.0859668477174502.

Tabel 12. Confusion Matrix

Aktual	Prediksi		
	Negatif	Netral	Positif
Negatif	202	0	8
Netral	23	0	2
Positif	112	0	25

Confusion matrix pada tabel 12 diatas dapat dijelaskan bahwa model mengklasifikasikan dengan benar sebesar 35 data positif, data netral 0, dan 337 data negatif. Sehingga, dari tabel confusion matrix diatas, maka diperoleh hasil klasifikasi dari Naive Bayes Classifier dengan perhitungan sebagai berikut:

$$Akurasi = \frac{227}{372} \times 100\% = 61\%$$

$$Precision\ Positif = \frac{25}{25+2+8} = 0.71$$

$$Precision\ Netral = \frac{0}{0+0} = 0$$

$$Precision\ Negatif = \frac{202}{202+23+112} = 0.60$$

$$Precision = \frac{0.71+0+0.60}{3} = 0.44 \times 100 = 44\%$$

$$Recall\ Positif = \frac{25}{112+0+25} = 0.18$$

$$Recall\ Netral = \frac{0}{23+0+2} = 0$$

$$Recall\ Negatif = \frac{202}{202+0+8} = 0.96$$

$$Recall = \frac{0.18+0+0.96}{3} = 0.38 \times 100 = 38\%$$

$$F1 - Score\ Positif = \frac{2 \times 0.18 \times 0.71}{0.18 + 0.71} = 0.29$$

$$F1 - Score\ Netral = \frac{2 \times 0 \times 0}{0+0+0} = 0$$

$$F1 - Score\ Negatif = \frac{2 \times 0.96 \times 0.60}{0.96 + 0.60} = 0.74$$

$$F1 - Score = \frac{2 \times 0.38 \times 0.44}{0.38 + 0.44} = 0.40 \times 100 = 40\%$$

KESIMPULAN

Proses pelabelan lexicon-based berhasil memberikan label terhadap dataset sebanyak 1860 tweet dengan rincian 1047 tweet bernilai negatif, 662 tweet bernilai positif, dan 151 tweet bernilai netral. Hasil dari proses lexicon-based digunakan untuk mengklasifikasi Naive Bayes dalam bentuk data uji hasil evaluasi klasifikasi. Sentiment pengguna twitter terhadap konflik perang Rusia Ukraina paling tinggi berada di tweet bersifat Negatif dengan metode dengan pembagian data training dan data testing yaitu 80:20 menunjukkan

metode Naïve Bayes Classifier dengan akurasi 61%, precision 44%, recall 38% dan f1-score 40%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada perusahaan XYZ, yang sudah bersedia menjadi objek penelitian mengenai pemilihan karyawan teladan. Penulis juga mengucapkan terima kasih kepada para pihak-pihak terkait yang sudah membantu dalam penelitian ini, sehingga bisa terlaksana dengan baik.

REFERENSI

- [1] A. Zaenudin, "Bagaimana Twitter Memengaruhi Opini Publik dan Preferensi Politik?" 21 Maret 2018. [Online]. Available: <https://tirto.id/bagaimana-twitter-memengaruhi-opini-publik-dan-preferensi-politik-cGre>.
- [2] Rizal, "Wow! Pengguna Aktif Harian Twitter Tembus 217 Juta di Dunia," 15 February 2022. [Online]. Available: <https://infokomputer.grid.id/read/123143849/wow-pengguna-aktif-harian-twitter-tembus-217-juta-di-dunia>.
- [3] A. Mardatila, "Penyebab Konflik Rusia dan Ukraina, Pahami Perkembangannya dari Waktu ke Waktu," merdeka.com, 24 Februari 2022. [Online]. Available: <https://www.merdeka.com/sumut/penyebab-konflik-rusia-dan-ukraina-pahami-perkembangannya-dari-waktu-ke-waktu-kln.html>.
- [4] A. Aflaha, "Rusia Vs Ukraina: Perang Akhirnya Beneran Terjadi!" cnbcindonesia, 24 February 2022. [Online]. Available: <https://www.cnbcindonesia.com/market/20224130324-17-318002/rusia-vs-ukraina-perang-akhirnya-beneran-terjadi>.
- [5] R. B. S. H. M. A. D. P. D. S. P. Haiqal Ramanizar Al Fajri, "Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan," Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), 2021.
- [6] H. N. P. M. W. d. I. Ali Sholihin, "nalysis Penyakit Difteri Berbasis Twitter Menggunakan," Sains, Aplikasi, Komputasi dan Teknologi Informasi, 2019.
- [7] K. I. Rifiana Arief, "ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN

- METODE LEXICON BASED," Jurnal Ilmiah MATRIK, 2019.
- [8] F. V. Sari, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID," Jurnal SIMETRIS, vol. 2, no. 681, p. November, 2019.
- [9] U. D. Arni, "Apa Itu Text Mining?" garudacyber, 25 Oktober 2018. [Online]. Available: <https://garudacyber.co.id/artikel/1254-apa-itu-text-mining>.
- [10] A. S. Oktriwina, "NLP: Kecerdasan Buatan yang Bantu Komputer Pahami Bahasa Manusia," glints, 2 Februari 2021. [Online]. Available: <https://glints.com/id/lowongan/natural-language-processing-adalah/#.YjEcWnpBxHM>.
- [11] D. Sierra, "Algoritma TF – IDF," medium, 13 Februari 2019. [Online]. Available: <https://dltsierra.medium.com/algoritma-tf-idf-633e17d10a80>.
- [12] I. S. d. D. F. Bitu Parga Zen, "TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News," Jurnal dan Penelitian Teknik Informatika, 2021.
- [13] N. S. Z. A. Dedi Darwis, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," Jurnal TEKNO KOMPAK, 2021.
- [14] F. V. Sari, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," Jurnal SIMETRIS, 2019.
- [15] G. A. B. M. B. S. Rachmad Mahendrajaya, "ANALISIS SENTIMEN PENGGUNA GOPAY MENGGUNAKAN METODE," Fakultas Teknik, Universitas Muhammadiyah Ponorogo, 2019.
- [16] R. Andriani and J. Attata, "UPAYA RESOLUSI KONFLIK RUSIA-UKRAINA," 28 March 2022.
- [17] R. N. Shobah, "Perang Rusia vs Ukraina: Sejarah Konflik 2 Negara Pecahan Uni Soviet Ini, Bermula dari Krisis 2013," Tribunkaltim.co, 25 Februari 2022. [Online]. Available: <https://kaltim.tribunnews.com/2022/02/25/perang-rusia-vs-ukraina-sejarah-konflik-2-negara-pecahan-uni-soviet-ini-bermula-dari-krisis-2013?page=2>.
- [18] sef, CNBC Indonesia, "Kronologi dan Latar Belakang Perang Rusia vs Ukraina," CNBC Indonesia, 6 Maret 2022. [Online]. Available: <https://www.cnbcindonesia.com/news/20220304133929-4-320041/kronologi-dan-latar-belakang-perang-rusia-vs-ukraina>.
- [19] D. H. Jayani, "10 Media Sosial yang Paling Sering Digunakan di Indonesia," 26 2 2020. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2020/02/26/10-media-sosial-yang-paling-sering-digunakan-di-indonesia>.
- [20] Y. Ratnasari, "Twitter Resmi Menambah Batasan Kicauan Jadi 280 Karakter," Tirto.id, 8 November 2017. [Online]. Available: <https://tirto.id/twitter-resmi-menambah-batasan-kicauan-jadi-280-karakter-czNm>.
- [21] F. D. Adhinata, B. Parga Zen and A. D. Septiadi, "Lightweight and Efficient Deep Learning Model for Fire Detection," 2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Palembang, Indonesia, 2023, pp. 583-588, doi: 10.1109/EECSI59885.2023.10295606.
- [22] F. Mukarromah and B. Parga Zen, "News Website Implementation Using The Fuzzy Analytical Hierarchy Process (FAHP) Method," 2023 29th International Conference on Telecommunications (ICT), Toba, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICT60153.2023.10374045.
- [22] A. Utami, D. F. Hebrasianto Permadi, B. P. Zen, F. D. Adhinata and W. A. Saputra, "Classification Analysis of Star Hotels in Banyumas Using the Random Forest Method," 2024 International Conference on Information Technology and Computing (ICITCOM), Yogyakarta, Indonesia, 2024, pp. 139-144, doi: 10.1109/ICITCOM62788.2024.10762547.
- [23] Bitu Parga Zen, Irwan Susanto, Khofifah Putriyani, Sintiya; Automatic document classification for tempo news articles about covid 19 based on term frequency, inverse document frequency (TF-IDF), and Vector Space Model (VSM). AIP Conf. Proc. 23 July 2024; 2952 (1): 060003. <https://doi.org/10.1063/5.0212036>
- [24] Zen, B.P., Abdurahman, Zafia, A., Utami, A., Putro, I.N.Y., Aditama, F.D. (2024). Multi Socket Transmission System Application with Advanced Encryption Standard

Algorithm to Support Confidential Medical Data Security. In: Triwiyanto, T., Rizal, A., Caesarendra, W. (eds) Proceedings of the 4th International Conference on Electronics, Biomedical Engineering, and Health Informatics. ICEBEHI 2023. Lecture Notes in Electrical Engineering, vol 1182. Springer, Singapore. https://doi.org/10.1007/978-981-97-1463-6_1.