ANALISIS SENTIMEN MASYARAKAT TERHADAP PENANGANAN KASUS PENEMBAKAN BRIGADIR J DENGAN ALGORITMA NAÏVE BAYES

Sinar Surya Prabu Al Amin, Jajam Haerul Jaman, Garno
Program Studi Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Karawang Jawa Barat
Sinar.surya18207@student.unsika.ac.id

ABSTRAK

Kepolisian adalah aparat pemerintahan yang bertugas untuk menjaga masyarakat indonesia dari kejahatan dan memberikan rasa nyaman dan aman untuk masyarakat indonesia. Pada penelitian bertujuan untuk mengetahui sentimental masyarakat terhadap kasus yang sempat viral di media sosial beberapa waktu yang lalu, untuk melakukan analisis sentimen masyarakat Indonesia terhadap penanganan kasus penembakan Brigadir Nofriansyah Yoshua Hutabarat (Brigadir J) oleh Irjen Pol Ferdy Sambo melalui media sosial Twitter. Maka dilakukan lah pengumpulan data dari tweet dengan menggunakan cara crawling data dari Twitter API, setelah mendapatkan data dari masyarakat melalui Twitter maka data akan di olah dengan beberapa cara yaitu pembersihan data dari noise dan karakter tidak relevan, terdapat 1720 data tweet yang digunakan dalam penelitian ini. Metode analisis sentimen yang digunakan adalah algoritma klasifikasi Naïve Bayes, dengan penerapan Knowledge Discovery in Database (KDD) yang meliputi Data Selection, Pre Processing, Transformation, Data Mining, dan Evaluation. Selanjutnya, dilakukan pembobotan sentimen menggunakan kamus lexicon dan kamus negative words untuk mengklasifikasikan data ke dalam tiga kategori: positif, negatif, dan netral. Hasil dari penelitian ini menunjukkan bahwa model dengan skenario pembagian data 70:30 memiliki kualitas analisis sentimen yang baik, dengan nilai akurasi sebesar 68% dan nilai AUC sebesar 0.76. Kata-kata yang sering muncul dalam dataset adalah "brigadir", "ferdy", "sambo", dan "bunuh". Penelitian ini memberikan wawasan tentang pandangan dan opini masyarakat terhadap penanganan kasus penembakan Brigadir J di Twitter. Diharapkan hasil penelitian ini dapat menjadi referensi data informasi mengenai opini masyarakat terkait isu sensitif dan membantu pihak berwenang dalam memahami persepsi publik terhadap kasus tersebut.

Kata kunci: Analisis sentimen, *Naive Bayes*, Brigadir, *Twitter*, *Sentimen skor*

1. PENDAHULUAN

Analisis sentimen ialah proses otomatis dalam memahami, mengekstrak, dan memproses data teks untuk menghasilkan informasi mengenai sentimen yang tersembunyi dalam sebuah pernyataan. Secara keseluruhan, tujuan dari analisis sentimen adalah untuk mengidentifikasi pandangan atau sikap yang diungkapkan oleh pembicara atau penulis terhadap suatu topik atau polaritas keseluruhan dalam suatu teks. Sikap ini bisa berwujud penilaian, evaluasi, ekspresi emosional penulis saat menulis, atau dampak emosional dari tulisan penulis pada para pembaca atau pendengar.

Twitter merupakan platform sosial yang interaktif yang memungkinkan pengguna untuk secara instan mengungkapkan kritik terhadap isu-isu secara waktu nyata.. Twitter yang banyak dimanfaatkan untuk menyuarakan pendapat. Dalam kurun lebih dari satu dekade, Twitter telah berhasil menarik perhatian dengan jumlah pengguna aktif bulanan mencapai 330 juta orang, dan setiap hari menerima sekitar 500 juta cuitan (tweet) (Twitter, 2019).

Salah satu topik yang sedang hangat dibicarakan hingga menjadi tren di platform Twitter adalah insiden penembakan yang menimpa Brigadir Nofriansyah Yoshua Hutabarat (Brigadir J) oleh Irjen Pol Ferdy Sambo. Pada kasus ini menjadi berkepanjangan dikarenakan skenario yang dibuat oleh Irjen Pol Ferdy Sambo yang merekayasa ulang kejadian pembunuhan tersebut. Dari penembakan yang dilakukan oleh Irjen Pol Ferdy Sambo kepada Brigadir J ini sudah melanggar HAM (Hak Asasi Manusia) dan melanggar pasal 340 KHUP tentang pembunuhan berencana dengan ancaman maksimal hukuman mati, penjara seumur hidup atau penjara selama-lamanya 20 tahun.

Kejadian ini pun menimbulkan komentar-komentar masyarakat di twitter dan bahkan menjadi trending topic hingga berhari-hari. Berbagai opini pun bermunculan terkait kasus penembakan brigadir J ini, baik positif, netral, maupun negatif yang mana bias saja mempengaruhi kepercayaan masyarakat terhadap kredibilitas Polri.

2. TINJAUAN PUSTAKA

2.1. Naïve Bayes Clasifier

Naïve Bayes Classifier diakui memiliki performa yang lebih unggul jika dibandingkan dengan beberapa metode klasifikasi lainnya seperti Decision Tree, Support Vector Machine, Fuzzy, dan sejenisnya. (Kurniawan, 2022). Algoritma naive bayes akan melakukan pekerjaan yang fantastis dalam mengklasifikasikannya. Rumus berikut mengilustrasikan teorema Bayes[1].

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Keterangan
X : Sampel data belum diketahui labelnya H : Hipotesis bahwa x merupakan data label
P(H) : Peluang dari hipotesa H
P(X) : Peluang data sampel
P(X|H) : Peluang data sampel apabila hipotesis benar
P(H|X) : Peluang hipotesis berdasarkan keadaan sampel.

Apabila X merujuk pada bukti, dan H adalah hipotesis, maka P(H | X) mewakili kemungkinan bahwa hipotesis H memiliki kebenaran berdasarkan bukti X, atau lebih tepatnya P(H | X) mewakili probabilitas posterior dari H dengan kondisi X. P(X | H) melambangkan probabilitas posterior dari X dengan asumsi H. Sementara itu, P(H) mencerminkan probabilitas awal hipotesis H, dan P(X) mengacu pada probabilitas awal bukti X..

2.2. *Crawling* Data

Crawling data di Twitter adalah tindakan mengakses informasi pengguna Twitter dan tweet dari server Twitter. Informasi ini diperoleh melalui Twitter API (API). Pengumpulan data ini atau langkah pengambilan data dari Twitter dilaksanakan melalui penerapan API Key Twitter, dengan bantuan bahasa pemrograman Python untuk mengelola prosesnya. API Key Twitter adalah sebuah Application Programming Interface (API) yang menyediakan beragam perintah, fungsi, komponen, serta protokol guna memfasilitasi para pengembang dalam membangun sistem perangkat lunak. Proses pengambilan data dilakukan melalui langkah pembuatan program yang melibatkan penggunaan kata kunci untuk mencari tweet sesuai dengan preferensi yang diinginkan[2].

2.3. Team Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) adalah representasi frekuensi kemunculan suatu kata, frasa, atau unit indeks lainnya dalam sebuah dokumen. Semakin jarang kata atau term muncul dalam berbagai dokumen, semakin tinggi nilai Inverse Document Frequency (IDF).[3].

$$tf - idf = tf \; x \; idf$$
$$idf = log \frac{n}{df}$$

Keterangan:
tf : *Term Frequency*
idf : *Inverse Document Frequency*
N : Jumlah total dokumen dalam corpus N
df : Jumlah dokumen yang mengandung term sebuah kata

2.4. Analisis Sentimen

Menurut Al Saeedi dan Khan dalam (Kurniawan, 2022) Analisis sentimen adalah metode untuk mengevaluasi pendapat yang diekspresikan dalam bentuk tulisan atau lisan dengan tujuan mengidentifikasi apakah pendapat tersebut bersifat positif, negatif, atau netral. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks dalam data, kalimat, atau fitur/tingkat aspek, serta menilai apakah pendapat yang disampaikan dalam data, kalimat, atau fitur entitas/aspek memiliki karakter positif, negatif, atau netral. Jika data tekstual semakin banyak yang dikumpulkan, maka akan semakin mudah untuk melakukan korelasi antar text dan jenis sentimen. Korelasi ini secara bersamaan dapat digunakan untuk memprediksi orientasi sentimen berbagai jenis teks[4].

2.5. *Text Mining*

Menurut Purbo dalam (Manik dkk., 2021) Text mining merupakan penerapan prinsip dan teknik data mining atau eksplorasi data pada dokumen yang memiliki bentuk teks dengan format yang tidak teratur atau beragam. Eksplorasi teks bertujuan untuk mengidentifikasi pola-pola berupa informasi dan pengetahuan yang memiliki nilai dalam konteks tertentu.[5].

2.6. Confusion Matrix

Menurut Yunita dalam (Anisa, 2019) onfusion matrix berfungsi sebagai perantara yang digunakan untuk menganalisis seberapa baik klasifikasi dapat mengenali elemen-elemen (tupel) dari kelas yang berbeda. Misalkan terdapat dua kelas, maka elemen-elemen tersebut akan diistilahkan sebagai tupel positif dan tupel negatif. True positive mengacu pada elemen-elemen dari kelas positif yang berhasil diidentifikasi dengan benar oleh sistem klasifikasi, sedangkan true negative adalah elemen-elemen dari kelas negatif yang juga diidentifikasi dengan tepat oleh sistem.
Namun, dalam klasifikasi terdapat kemungkinan terjadinya kesalahan. False positive terjadi ketika elemen-elemen dari kelas negatif salah diidentifikasi sebagai elemen-elemen kelas positif, sementara false negative terjadi ketika elemen-elemen dari kelas positif salah diidentifikasi sebagai elemen-elemen kelas negatif[6].

2.7. Data Mining

Menurut Larose dalam[7]. Data mining merupakan langkah dalam menemukan makna melalui identifikasi pola dalam sekumpulan data besar. Proses ini melibatkan penyaringan data yang tersimpan dalam suatu repositori, menggunakan statistik dan teknik matematika, dengan tujuan mengungkapkan informasi yang memiliki signifikansi.
Data Mining adalah proses mengorek data yang terkumpul untuk mengubahnya menjadi informasi berharga bagi penggunanya. Ada beberapa teknik yang dikenal dan umum digunakan dalam data mining, antara lain klasterisasi, klasifikasi, asosiasi, dan teknik-teknik lain yang berkembang sesuai dengan perubahan tren data saat ini.

2.8. KDD (Kmowledge Discovery in Database)

Menurut Liliana Swastina dalam (Anisa, 2019) *Knowledge Discovery In Databases (KDD)* merujuk pada proses yang tidak sederhana untuk menemukan dan mengenali pola dalam data. Pola yang ditemukan harus memiliki kriteria validitas, kebaruan, kemanfaatan, dan pemahaman. *KDD* melibatkan integrasi teknik dan penemuan ilmiah, serta interpretasi dan visualisasi dari pola-pola yang ditemukan dalam sejumlah kumpulan data. *KDD* secara umum terdiri dari lima tahap, yaitu pemilihan data, pra-pemrosesan/pembersihan, transformasi, penambangan data, dan interpretasi/evaluasi. [8]

2.9. Twitter API

Twitter API (Application Programming Interface) merupakan fasilitas yang disediakan oleh Twitter untuk melakukan pengumpulan data. API Twitter sendiri adalah tools yang paling baik untuk mengumpulkan data yang dihasilkan melalui interaksi antar pengguna Twitter. Untuk dapat mengakses API, pengguna dapat mengajukan akun pengembang atau developer. Setelah aplikasi disetujui, maka pengguna memiliki akses atas 4 macam kunci yaitu consumer key, consumer secret, access token, dan access secret. Kunci-kunci tersebut nantinya dapat mengotentikasi pengguna untuk mengakses data Twitter[9].

2.10. Twitter

Twitter merupakan suatu platform media sosial yang sangat diminati oleh pengguna internet, khususnya kalangan muda. Hal ini disebabkan oleh kemampuan Twitter yang memungkinkan pengguna untuk berbagi pendapat atau opini, yang sering disebut sebagai "cuitan". Melalui tulisan di Twitter, individu dapat dengan bebas berbagi informasi dan mengemukakan pandangan tanpa batasan.

Dalam kurun waktu lebih dari satu dekade sejak didirikan, Twitter berhasil mencapai angka 330 juta pengguna aktif setiap bulan, dengan jumlah cuitan (tweet) yang mencapai sekitar 500 juta setiap harinya (Twitter, 2019). Jadi sangat mungkin untuk melakukan suatu survei sebuah opini karena dengan banyak nya pengguna aktif, salah satunya di Indonesia[10].

3. METODE PENELITIAN

Metodologi penelitian yang diterapkan dalam studi ini ialah knowledge discovery in Database (KDD). Proses tahapan metodologi ini dengan mengadopsi metode KDD melibatkan langkah-langkah seperti pemilihan data, pra-pemrosesan data, transformasi data, penambangan data, serta evaluasi dan interpretasi pengetahuan. Adapun tahapan KDD yang digunakan yaitu:

1. *Data Selection*
   Pada tahap awal ini, data-data pada database akan diseleksi kemudian disimpan didalam berkas yang terpisah.
2. *Pre-processing/Cleaning*
   Tahapan berikutnya adalah melakukan pembersihan terhadap data-data yang sudah diperoleh sebelum dilakukan proses data mining.
3. *Transformation*
   Tahapan ini merupakan suatu proses dimana akan dilakukannya transformasi atau perubahan bentuk data terhadap data-data yang sudah di bersihkan sehingga nantinya data tersebut dapat dilanjutkan ke dalam proses data mining.
4. *Data Mining*
   Di dalam tahapan ini akan dilakukan proses pencarian pola atau informasi dengan menggunakan metode atau teknik tertentu terhadap data-data.
5. *Interpretation/Evaluation*
   Tahap ini merupakan tahap terakhir yang dimana akan dilakukannya proses evaluasi terhadap hasil dari proses data mining yang sudah dilakukan.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah mengetahui sentimen dari cuitan-cuitan para pengguna Twitter di Indonesia terkait kasus penembakan brigadir JTweet-tweet yang ada akan diurutkan menjadi tiga kelompok berbeda, yaitu positif, negatif, dan netral, dengan bantuan metode klasifikasi Naive Bayes Classifier. Hasil klasifikasi dan perhitungan skor sentimen dapat menjadi referensi data informasi tentang pandangan atau opini masyarakat terkait kasus tersebut di masa mendatang.

4.1. Data Selection

Hasil dari langkah ini menghasilkan himpunan data yang akan dipergunakan dalam penelitian ini. Metode yang diterapkan adalah pengambilan data (crawling) melalui Twitter API pada tanggal 2 November 2022.

	username	tweetcreatedts	text
0	AyuWulandariz	2022-11-24 12:14:16	4 Kesamaan dari Kematian Tak Wajar Prada Indra...
1	M453h47	2022-11-24 12:11:59	Kamaruddin Simanjuntak : "Clin² Psikopat ada pa...
2	tribunnews	2022-11-24 12:10:00	Ridwan Soplanit Bersaksi Ferdy Sambo Ikut Temb...
3	didienAZHAR	2022-11-24 12:09:02	Mirip Kasus Brigadir J, Alasan TNI AU Berikan ...
4	alvi_haddim	2022-11-24 12:06:41	Upaya Lindungi Putri Candrawathi, Saor Siagian...
...	...	...	...
5064	lilsunast	2022-11-02 05:03:26	💙 antara kesel & greget ntn sidang ini, ma...
5065	inilahdotcom	2022-11-02 05:03:06	Kuasa Hukum terdakwa Putri Candrawathi, Sarmau...
5066	purba_keykelly	2022-11-02 05:01:40	Rosti Simanjuntak ibu almarhum Yosua Hutabarat...
5067	jayalperkasa	2022-11-02 05:01:16	Ayah brigadir J minta kepada majelis hakim, ag...
5068	CNNIndonesia	2022-11-02 05:01:07	Foto Jenazah Brigadir J Ditampilkan di Persida...

5069 rows × 3 columns

Gambar 1. Dataset Penelitian dari Hasil Selection

Data yang diperoleh berupa kumpulan cuitan dengan jangka waktu yang dimulai dari tanggal... 2 November 2022 hingga 30 November 2022, dan berhasil mengumpulkan 5.068 tweet, seperti yang terlihat pada Gambar 4.1.

4.2. *Pre Proccesing/ Cleaning*

Setelah pengumpulan kumpulan data, terungkap bahwa jenis dataset yang ada belum optimal untuk keperluan proses data mining. Maka dari itu, langkah

pra-pemrosesan perlu diterapkan untuk menghapus kata-kata atau karakter yang tidak relevan.

Dalam hasil contoh tersebut, sebelum mengalami tahap pembersihan (cleaning), masih terdapat URL, karakter, emotikon, dan hashtag yang tidak relevan untuk proses klasifikasi. Tetapi, setelah melewati tahap pembersihan, tweet tersebut telah dibersihkan sehingga dapat dilakukan proses klasifikasi dengan baik.

Pada langkah ini, terjadi transformasi semua huruf besar dalam dataset menjadi huruf kecil. Proses ini... bertujuan untuk melakukan penyeragaman karakter dalam data yang akan diproses, sehingga pada tahap preprocessing selanjutnya dapat Menghasilkan hasil pra-pemrosesan yang efektif. Di bawah ini merupakan ilustrasi hasil tahap penyederhanaan huruf (case folding) yang terdapat dalam Tabel 2.

Tabel 1. Tabel contoh hasil data cleaning

Twit sebelum dilakukan cleaning	**Tweet setelah dilakukan cleaning**
Belum ada temuan oleh Timsus Polri terkait sosok kakak asuh Ferdy Sambo yang disebut meringankan vonisnya pada kasus Brigadir J. #TempoNasional https://t.co/yHaTGJ0EIm	Belum ada temuan oleh Timsus Polri terkait sosok kakak asuh Ferdy Sambo yang disebut meringankan vonisnya pada kasus Brigadir J
Sangat Menyayangkan Sikap Jokowi, Pengacara Brigadir J Membicarakan Pilpres 2024, Ajak Masyarakat Â Pilihâ€ https://t.co/uHYz7uShnL	Sangat Kecewa dengan Jokowi, Pengacara Brigadir J Ungkap Isu Pilpres 2024 dan Mengajak Masyarakat untuk Memilih

Tabel 2. Tabel contoh hasil tahap *case folding*

Twit sebelum dilakukan Case Folding	**Tweet setelah dilakukan Case Folding**
Ketidak berpihakan Irjen Ferdy Sambo Terhadap Ucapan Bharada E Brigadir J yang Sudah Menyerah dan Ketakutan, Namun Tetap Mengalami Akhir yang Tragis Meskipun Telah Memohon Maaf.	Tindakan kejam Irjen Ferdy Sambo yang diutarakan oleh Bharada E Brigadir J, yang telah menyerah dan merasa takut, tetap berujung pada penghapusan meskipun permohonan maaf sudah diajukan.
Sosok yg mampu mendengarkan dan yakin dgn nuraninya Sebaliknya Bharada E justru menyelamatkan diri dan dgn teganya menembak mati sahabatnya sendiri Sungguh memilukan	sosok yg mampu mendengarkan dan yakin dgn nuraninya sebaliknya bharada e justru menyelamatkan diri dan dgn teganya menembak mati sahabatnya sendiri sungguh memilukan
Komnas HAM menyebut ada penembak lain selain Tersangka FS dan Bharada E pernyataan ini disampaikan kemungkinan untuk mengimbangi rekomendasi dari Komnas HAM bahwa adanya Pelecehan Seksual di Magelang karena Komnas HAM di Buli	komnas ham menyebut ada penembak lain selain tersangka fs dan bharada e pernyataan ini disampaikan kemungkinan untuk mengimbangi rekomendasi dari komnas ham bahwa adanya pelecehan seksual di magelang karena komnas ham di buli

Setelah melakukan tahap case folding, terlihat bahwa semua huruf dalam dataset berubah menjadi huruf kecil, termasuk kata-kata yang seharusnya dalam huruf kapital. Proses ini bertujuan untuk mempermudah proses klasifikasi. Sebagai contohnya, kata "HAM" yang sebelumnya ditulis dalam huruf kapital, setelah mengalami tahap case folding berubah menjadi huruf kecil "ham".

Setelah membersihkan data dan mengubahnya menjadi huruf kecil, langkah berikutnya adalah tokenizing. Tokenizing adalah langkah pembagian kalimat dalam data komentar menjadi kata per kata. Proses ini akan memberi kemudahan saat beralih ke tahap transformasi, karena data akan diproses per kata daripada per kalimat. Hasil dari proses tokenizing dapat dilihat pada Tabel 3.

Tabel 3. Tabel hasil proses *tokenizing*

Twit sebelum dilakukan Tokenizing	**Tweet setelah dilakukan Tokenizing**
Acara Brigadir J Kebanyakan Lebar Omong Fokus Aja Brigadir J Udah Tau Kapolri Udah Pecat Sambo Banyak Bacot Acara	‘acara’,‘brigadirj’, ‘kebanyakan’, ‘lebar’, ‘omong’, ‘fokus’, ‘aja’, ‘brigadirj’, ‘udah’, ‘tau’, ‘kapolri’, ‘udah’, ‘pecat’, ‘sambo’,‘banyak’,‘bacot’, ‘acara’

Semangat Kamaruddin BiarpunSelesai Bela Keluarga Brigadir J Tetap Suara Benar Jujur	‘semangat’, ‘kam aruddin’, ‘biarpun’, ‘selesai’, ‘bela’, ‘keluarga’, ‘brigadirj’, ‘tetap’, ‘suara’, ‘benar’, ‘jujur’

Pada Tabel 3. memperlihatkan hasil dari Pada tahap tokenisasi, data cuitan yang sebelumnya berupa rangkaian kalimat telah berhasil diuraikan menjadi kata-kata per kata. Setiap kata dipisahkan oleh tanda kutip satu yang diletakkan di awal dan akhir huruf pada setiap kata.

Langkah selanjutnya adalah tahap Stopword Removal, di mana kata-kata yang tidak relevan akan disaring untuk proses klasifikasi lebih lanjut. Contoh kata-kata yang tak relevan adalah "yang," "dan," "di," "dari," serta "nya." Dalam tahap ini, akan memanfaatkan dokumen yang berisi daftar kata stopword untuk membantu menyaring kata-kata yang tidak bermanfaat dalam proses klasifikasi. Output dari tahap Penghapusan Stopword yang didukung oleh dokumen yang berisi daftar kata stopword dapat diamati pada Tabel 4.

Tabel 4. Hasil proses filtering atau *stopward removal*

Twit sebelum dilakukan Filtering	**Tweet setelah dilakukan Filtering**
‘gw’,‘denger’,‘ulang’, ‘omongannya’, ‘si’, ‘gl’, ‘ke’,‘bapaknya’,‘alm’,‘brigadir’,‘j’,‘lgsg’,‘emosi’,‘dan’,‘amarah’, ‘gw’,‘naik’, ‘jd’, ‘ga’,‘bs’, ‘tidur’, ‘setan’, ‘alas’, ‘itu’, ‘pendeta’,‘masuk’,‘neraka’,‘aja’, ‘lu’, ‘keparat’	gw’,‘denger’, ‘ulang’, ‘omongannya’, ‘gl’, ‘ke’, ‘alm’,‘brigadir’, ‘j’, ‘lgsg’, ‘emosi’,‘amarah’, ‘gw’, ‘jd’, ‘ga’, ‘bs’,‘tidur’, ‘setan’, ‘alas’, ‘pendeta’, ‘masuk’, ‘neraka’, ‘aja’, ‘lu’, ‘keparat’
‘rakyat’, ‘kdu’,‘kawal’, ‘pembunuh’,‘brigadir’,‘j’, ‘agar’, ‘tersangka’, ‘bsa’, ‘di’,‘jerat’, ‘hukum’, ‘yg’, ‘sesuai’	‘rakyat’, ‘kdu’,‘kawal’, ‘bunuh’,‘brigadir’,‘sangka’, ‘bsa’, ‘jerat’, ‘hukum’, ‘sesuai’

Pada contoh diatas dijelaskan bahwa ada beberapa kata yang dihilangkan seperti kata terus, lanjut dan sebagainya.

Langkah akhir dalam pra-pemrosesan adalah proses stemming, yang bertujuan untuk mengubah kata-kata dalam dokumen menjadi bentuk dasarnya. Proses ini mengaplikasikan pustaka kata dasar. Perbedaan antara kondisi sebelum dan setelah proses stemming ditampilkan pada Tabel 5.

Tabel 5. Proses kata sebelum dan sesudah *stemming*

Kata Sebelum *Stemming*	**Kata Sesudah *Stemming***
Pembunuh	Bunuh
Penembakan	Tembak
Kekuatan	Kuat
Kekuasaan	Kuasa
Menggunakan	Guna
Bantuaan	Bantu

Pada tabel 5. Terdapat contoh kata kata yang sudah melalui proses stemming, dan hasil proses stemming di dalam dataset dapat dilihat pada tabel 6.

Tabel 6. Hasil proses *stemming*

Twit sebelum dilakukan Stemming	**Tweet setelah dilakukan Stemming**
‘gw’,‘denger’,‘ulang’, ‘omongannya’,‘si’,‘gl’,‘ke’,‘bapaknya’,‘alm’,‘brigadir’,‘j’,‘lgsg’,‘emosi’,‘dan’,‘amarah’, ‘gw’, ‘naik’, ‘jd’, ‘ga’, ‘bs’, ‘tidur’, ‘setan’, ‘alas’,‘itu’,‘pendeta’,‘masuk’, ‘neraka’, ‘aja’, ‘lu’, ‘keparat’	‘denger’,‘ulang’,‘omongan’ ‘brigadir’,‘emosi’,‘amarah’ ‘tidur’, ‘setan’, ‘alas’, ‘pendeta’,‘masuk’, ‘neraka’, ‘aja’, ‘keparat’
‘rakyat’,‘kdu’,‘kawal’, ‘pembunuh’,‘brigadir’,‘j’,‘agar’, ‘tersangka’, ‘bsa’,‘di’‘jerat’, ‘hukum’, ‘yg’, ‘sesuai’	‘rakyat’, ‘kawal’, ‘bunuh’, ‘brigadir’,‘sangka’,‘jerat’, ‘hukum’, ‘sesuai’

Dilihat dari Tabel 6, kata 'pembunuh' telah mengalami proses stemming dengan menghapus imbuhan yang ada sehingga kembali menjadi bentuk kata dasarnya yaitu 'bunuh'. Setelah proses preprocessing dilakukan, jumlah dataset berkurang menjadi 1720 karena beberapa data yang dihilangkan tidak relevan untuk proses klasifikasi dan juga terdapat banyak duplikasi data

4.3. *Transformation Data*

Tahap selanjutnya setelah selesai melakukan tahap preproccesing, yaitu Langkah transformasi data. Sebelum menerapkan transformasi data, langkah awal dilakukan dengan melakukan pembobotan sentimen untuk memberikan label pada data. Dalam tahap pembobotan sentimen, dilakukan perhitungan skor sentimen dengan menghitung jumlah kata yang memiliki sentimen positif dan negatif. Skor ini dihasilkan melalui analisis polaritas yang terdapat pada dataset. Pelabelan data kata bersentimen positif dan kata bersentimen negatif dibantu dengan kamus bahasa yang sudah di tersedia.

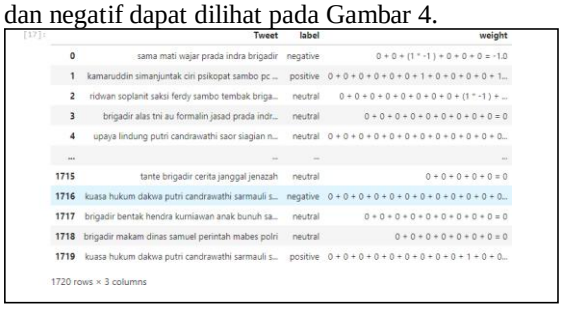
Kamus sentimen dan negatif tersebut di dapatkan dengan mengunduh dokumen khusus yang berisi kalimat yang berbobot positif dan kalimat berbobot negatif. Untuk dokumen yang berisi kamus sentimen positif dan negatif dapat dilihat pada lampiran 4. Berikut contoh isi dari dokumen khusus yang berisi kamus kalimat berbobot positif, dapat dilihat pada gambar 2.

Gambar 2. Kamus Sentimen positif

Pada gambar 2 pembobotan sentimen dihitung dengan memberi nilai, untuk kalimat sentimen berbobot positif diberi nilai 1 dan untuk kalimat sentimen berbobot negatif diberi nilai -1. Dan Berikut contoh isi dari dokumen khusus yang berisi kamus kalimat berbobot negatif, dapat dilihat pada Gambar 3.

Gambar 3. Kamus sentimen negatif

Setelah menambahkan kamus skor sentimen positif dan negatif dengan mengunduh dokumen khusus, Oleh karena itu, langkah berikutnya adalah melakukan perhitungan skor sentimen dengan mengevaluasi polaritas dari setiap kalimat yang terdapat dalam data cuitan. Apabila skor sentimen tersebut memiliki nilai yang lebih besar dari nol, maka sentimen tersebut akan dikategorikan sebagai positif. Sebaliknya, jika skor sentimen memiliki nilai yang kurang dari nol, maka kalimat tersebut akan masuk ke dalam kategori negatif. Hasil pembobotan skor sentimen dengan menggunakan kamus kalimat positif dan negatif dapat dilihat pada Gambar 4.



	Tweet	label	weight
0	sama mati wajar prada indra brigadir	negative	0 + 0 + (1 * -1) + 0 + 0 + 0 = -1.0
1	kamaruddin simanjuntak ciri psikopat sambo pc ...	positive	0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 1...
2	ridwan soplanit saksi ferdy sambo tembak briga...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + (1 * -1) + ...
3	brigadir alas tni au formalin jasad prada indr...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
4	upaya lindung putri candrawathi saor siagian n...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
...	...	...	...
1715	tante brigadir cerita janggal jenazah	neutral	0 + 0 + 0 + 0 + 0 + 0 = 0
1716	kuasa hukum dakwa putri candrawathi sarmauli s...	negative	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0...
1717	brigadir bentak hendra kurniawan anak bunuh sa...	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
1718	brigadir makam dinas samuel perintah mabes polri	neutral	0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
1719	kuasa hukum dakwa putri candrawathi sarmauli s...	positive	0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0...

1720 rows × 3 columns

Gambar 4. Hasil pembobolan skor sentimen

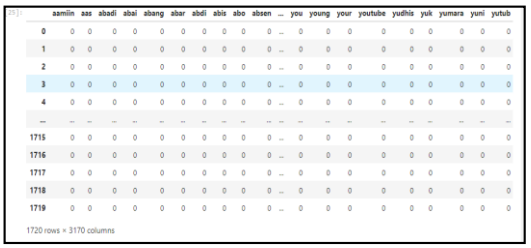
Hasil dari pembobotan skor sentimen dilanjut kan ketahap klasifikasi selanjutnya dengan menyimpan ke dalam file berformat CSV. Ringkasan hasil dari jumlah pelabelan data sentimen positif dan negatif, dapat dilihat pada Tabel 7.

Tabel 7. Jumlah skor sentimen pelabelan data.

Label/ Kelas	Jumlah
Positif	325
Negatif	301
Netral	1094
Total	**1720**

Hasil pembobotan skor sentimen pelabelan data berjumlah 1720 data dengan kelas positif berjumlah 325 data, kelas negatif 301 data, dan kelas netral berjumlah 1094. Setelah melakukan pembobotan skor sentimen dan pelabelan data dilanjutkan ke proses transformation data.

Pada tahap ini, dilakukan pemberian nilai bobot pada kata-kata menggunakan metode pembobotan TF (*Term Frequency*) dari data komentar yang telah melalui langkah pelabelan dan pembobotan skor sentimen. Berikut adalah hasil dari proses transformasi data, yang dapat ditemukan dalam ilustrasi pada gambar 5.



	aamiin	aas	abadi	abai	abang	abar	abdi	abis	abo	absen	...	you	young	your	youtube	yudhis	yuk	yumara	yuni	yutub
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1715	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1716	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1717	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1718	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1719	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0

1720 rows × 3170 columns

Gambar 5. Hasil proses transformation data melalui implementasi TF-IDF

Pada Gambar 5 dapat dilihat pada setiap kata/terms dalam corpus mempunyai bobot nilainya masing – masing dan bisa dilanjutkan ke tahap klasifikasi selanjutnya.

4.4. *Data Mining*

Setelah melalui langkah transformasi data, langkah selanjutnya adalah proses penambangan data. Tahap ini merupakan langkah klasifikasi data yang dijalankan dengan menerapkan model dari algoritma Naive Bayes, yakni Multinomial Naive Bayes, dengan memanfaatkan pustaka sklearn. naive_bayes yang tersedia dalam bahasa pemrograman Python. Ilustrasi implementasi algoritma Multinomial Naive Bayes dapat ditemukan dalam Gambar 6 sebagai berikut.

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB().fit(X_train_df, y_train)
prediction_mi = model.predict(X_test_df)
prediction_proba_mi = model.predict(X_test_df)
```

Gambar 6. Library Multinominal Bayes.

Dalam riset ini, metode pengolahan data yang digunakan menggabungkan algoritma Naïve Bayes dalam empat skenario berbeda, yaitu rasio 90:10, 80:20, 70:30, dan 60:40. Penilaian sentimen terkait penanganan kasus Brigadir J ini menggunakan Naïve Bayes Classifier sebagai alat klasifikasi, dengan Confusion Matrix berperan dalam memberikan nilai akurasi prediksi mesin terhadap data yang telah diolah. Berikut adalah pembagian antara data latihan (training) dan data uji (testing)

Tabel 8. Tabel skenario pembagian Split Data

Presentase Data		Jumlah Data	
Training	**Testing**	**Training**	**Testing**
90%	10%	1548	172
80%	20%	1376	344
70%	30%	1204	516
60%	40%	1032	688
Total		**1720**	

4.5. *Evaluation*

Setelah tahap pembentukan model berhasil diselesaikan, model yang dihasilkan dari tiap skenario akan dites. Evaluasi hasil pengujian model ini meliputi nilai akurasi, presisi, recall, dan f-measure (f1-score). Berikut ini merupakan perhitungan evaluasi untuk setiap model skenario yang telah diuji.

1. Skenario 1 (90% : 10%)

Ilustrasi di bawah ini menunjukkan hasil klasifikasi yang dilakukan menggunakan komposisi 90% data latihan dan 10% data uji.

```
                    actual:negatif  actual:netral  actual:positif
predicted:negatif                6             11               7
predicted:netral                17             88               7
predicted:positif                9             13              14

Multinomial NB Accuracy  :  0.627906976744186
Multinomial NB Precision :  0.4748677248677248
Multinomial NB Recall    :  0.49107142857142855
Multinomial NB F-Measure :  0.4791666666666667
```

Gambar 7. Hasil skenario 1

Berdasarkan gambar 7, Dari klasifikasi yang menerapkan metode Naive Bayes pada rasio 90:10, tercatat bahwa hasilnya menghasilkan tingkat akurasi sebesar 62%, presisi sebesar 47%, recall sebesar 49%, dan nilai f-measure sebesar 47%.

2. Skenario 2 (80% : 20%)

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 80% data training dan 20% data testing.

```
                    actual:negatif  actual:netral  actual:positif
predicted:negatif                8             13               9
predicted:netral                29            189              17
predicted:positif               21             27              31

Multinomial NB Accuracy  :  0.6627906976744186
Multinomial NB Precision :  0.4877756830355808
Multinomial NB Recall    :  0.5023727315075321
Multinomial NB F-Measure :  0.48411856905771716
```

Gambar 8. Hasil skenario 2.

3. Skenario 3 (70% - 30%)

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 70% data training dan 30% data testing

```
                    actual:negatif  actual:netral  actual:positif
predicted:negatif               13             18               8
predicted:netral                46            283              24
predicted:positif               30             38              56

Multinomial NB Accuracy  :  0.6821705426356589
Multinomial NB Precision :  0.5288819844243403
Multinomial NB Recall    :  0.5390797705603313
Multinomial NB F-Measure :  0.5164486539789871
```

Gambar 9 Hasil skenario 3

Berdasarkan gambar 9, Dari hasil klasifikasi yang menggunakan algoritma Naive Bayes dengan rasio 70:30, diperoleh nilai akurasi sebesar 68%, presisi sebesar 52%, recall sebesar 53%, dan nilai f-measure sebesar 51%.

4. Skenario 4 (60% : 40%)

Gambar dibawah ini adalah hasil dari klasifikasi dengan menggunakan persentase 60% data training dan 40% data testing

```
                    actual:negatif  actual:netral  actual:positif
predicted:negatif               17             28              12
predicted:netral                57            370              38
predicted:positif               46             49              71

Multinomial NB Accuracy  :  0.6656976744186046
Multinomial NB Precision :  0.5072184607132548
Multinomial NB Recall    :  0.5187280061136071
Multinomial NB F-Measure :  0.4994224744719729
```

Gambar 10. Hasil skenario 4

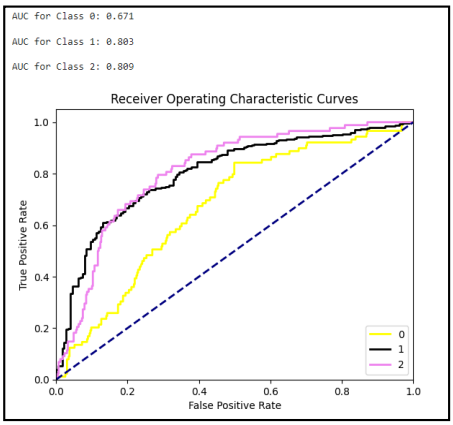
Berdasarkan illustrasi pada Gambar 10, hasil dari klasifikasi yang mengadopsi algoritma Naive Bayes dengan rasio 60:40 menghasilkan akurasi sebesar 66%, presisi sebesar 50%, recall sebesar 51%, dan f-measure sebesar 49%.

Selain menggunakan matriks konfusi untuk mengevaluasi kinerja model, nilai Area Under Curve (AUC) dan kurva Receiver Operating Characteristics (ROC) juga dapat digunakan untuk mengukur performa model. Nilai AUC dari seluruh model Multinomial Naive Bayes tertera pada Tabel 9 berikut.

Tabel 9. Hasil nilai AUC setiap skenario

Skenario	AUC	Kualitas
90:10	0.76	*Fair Classification*
80:20	0.76	*Fair Classification*
70:30	0.76	*Fair Classification*
60:40	0.75	*Fair Classification*

Dari tabel di atas, semua model dalam berbagai skenario menghasilkan nilai AUC yang menunjukkan kualitas klasifikasi yang bersifat cukup baik. Nilai AUC terendah terdapat pada model dalam skenario 60:40 dengan nilai 0.75, sementara nilai AUC pada model dalam skenario 90:10, 80:20, dan 70:30 mempunyai nilai yang sama yaitu 0.76. Berikut grafik ROC dari model 70:30



Gambar 11. Grafik ROC skenario 70:30\

Pada Gambar 11, diberikan presentasi nilai AUC untuk kelas positif, yang ditunjukkan dengan garis berwarna kuning, dengan nilai 0.671. Selanjutnya, untuk kelas netral, ditunjukkan dengan garis hitam dan memiliki nilai AUC sebesar 0.803, sementara kelas negatif, yang ditunjukkan oleh garis merah muda, memiliki nilai AUC sebesar 0.809. Dengan mempertimbangkan nilai-nilai ini pada tiap kelas, maka dapat dihasilkan nilai ROC AUC score pada model yaitu:

$$ROC\ AUC\ Score = \frac{0.671 + 0.803 + 0.809}{3} = 0.76$$

Dapat diartikan bahwa model skenario 70:30 dengan perolehan nilai AUC yaitu 0.76 mempunyai kualitas Fair classification.

5. KESIMPULAN DAN SARAN.

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, yakni: Hasil penelitian ini menunjukkan sentimen cuitan para pengguna Twitter di Indonesia terkait kasus penembakan brigadir J. Cuitan-cuitan kemudian dianalisis dan dikelompokkan ke dalam tiga kategori, yakni positif, negatif, dan netral, dengan bantuan metode klasifikasi Naive Bayes Classifier. Total data dalam penelitian ini berjumlah 1720, dengan kelas positif memiliki 325 data, kelas negatif 301 data, dan kelas netral 1094 data. Hasilnya menunjukkan bahwa sentimen netral memiliki jumlah data yang paling banyak, menunjukkan bahwa mayoritas cuitan terkait kasus penembakan brigadir J termasuk ke dalam kategori netral. Dari hasil data yang menampilkan sentimen netral mendapatkan persentase paling besar daripada sentimen positif dan negatif, menandakan bahwa tingkat kepercayaan masyarakat terhadap penanganan kasus penembakan brigadir J ini termasuk ke dalam kategori netral.

Dari klasifikasi yang menerapkan metode Naive Bayes pada rasio 90:10, tercatat bahwa hasilnya menghasilkan tingkat akurasi sebesar 62%, presisi sebesar 47%, recall sebesar 49%, dan nilai f-measure sebesar 47%.

Dari hasil klasifikasi yang menggunakan algoritma Naive Bayes dengan rasio 70:30, diperoleh nilai akurasi sebesar 68%, presisi sebesar 52%, recall sebesar 53%, dan nilai f-measure sebesar 51%.
Dari hasil klasifikasi yang mengadopsi algoritma Naive Bayes dengan rasio 60:40 menghasilkan akurasi sebesar 66%, presisi sebesar 50%, recall sebesar 51%, dan f-measure sebesar 49%.

DAFTAR PUSTAKA

[1] F. Darwis *et al.*, "SISTEM ANALISIS SENTIMEN PUBLIK TENTANG OPINI PEMILIHAN KEPALA DAERAH JAWA TIMUR 2018 PADA DOKUMEN TWITTER MENGGUNAKAN NAIVE BAYES," 2018.

[2] J. E. Sembodo, E. B. Setiawan, dan Z. A. Baizal, "Data Crawling Otomatis pada Twitter," in *Indonesian Symposium on Computing (Indo-SC)*, 2016, hal. 11–16.

[3] R. Melita *et al.*, "(TF-IDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS : SYARAH UMDATIL AHKAM)," vol. 11, no. 2, 2018.

[4] A. Sentimen dan F. Media, "ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP PRODUK INDIHOME DAN FIRST MEDIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK REVIEW ANALYSIS SENTIMENT CUSTOMER PRODUCT INDIHOME AND FIRST MEDIA USING CONVOLUTIONAL NEURAL NETWORK," vol. 8, no. 5, hal. 9049–9061, 2021.

[5] M. W. Berry dan M. Castellanos, "Survey of Text Mining : Clustering , Classification , and Retrieval , Second Edition," 2007.

[6] M. M. Degree, C. Science, dan A. C. Lecture, "Data Mining : Concepts and," 2012.

[7] J. Han, J. Pei, dan M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[8] A. Novantirani, M. K. Sabariah, dan V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeeding Eng.*, vol. 2, no. 1, hal. 1–7, 2015.

[9] W. Gata dan P. P. Purnomo, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," *Format*, vol. 6, no. 1, hal. 1–13, 2017.

[10] I. Ali, M. Asif, I. Hamid, dan M. Umer, "A word embedding technique for sentiment analysis of social media to understand the relationship between Islamophobic incidents and media portrayal of Muslim communities," no. January, 2022, doi: 10.7717/peerj-cs.838.