

ANALISIS SENTIMEN GENERASI Z TERHADAP PENGETAHUAN TENTANG KEHIDUPAN GAY MENGGUNAKAN ALGORITMA NAÏVE BAYES

Peni¹, Nana Suarna², Willy Prihartono³

¹Program Studi Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Indonesia

²Program Studi Sistem Informasi, STMIK LIKMI Bandung

Jl. Ir. H. Djuanda No. 96, Bandung, Jawa Barat, Indonesia

³Program Studi Sistem Informasi, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Indonesia

peni252525@gmail.com

ABSTRAK

Dalam perkembangan zaman dan semakin majunya teknologi informasi yang berkembang setiap saat memudahkan generasi z untuk mendapatkan berbagai kemudahan khususnya pada penggunaan media sosial seperti twitter, permasalahan pada penelitian ini untuk mengetahui Berapa nilai akurasi, presisi, dan recall pada analisis sentimen pada kehidupan gay berupa ciri gay, Metode yang digunakan dalam penelitian ini menggunakan metode alur penelitian *Knowledge Discover in Database* atau (KDD). pada penelitian ini peneliti melakukan pengumpulan data pengguna media sosial twitter melalui proses scraping data pada twitter pada akun @masndutx tentang ciri gay. Data yang didapatkan pada penelitian ini sebanyak 1500 data namun setelah dilakukan pelabelan data dan proses *cleansing* data yang digunakan sebanyak 1010 data teks. Berdasarkan hasil penelitian data teks analisis sentimen generasi z terhadap pengetahuan gay yang menggunakan algoritma naïve bayes menggunakan (KDD) Keakuratannya diuji menggunakan 1010 data dari pengujian yang didapatkan generasi z yang mendukung tentang gay dan yang tidak mendukung gay pada twitter adalah mereka yang setuju atau mendukung dengan pendapat tentang gay dan ciri gay pada twitter, tidak sedikit yang memberikan pendapatnya tidak setuju dengan ciri-ciri gay yang ada pada twitter Dan hasil pengujian klasifikasi menggunakan metode Algoritma naïve bayes didapatkan *accuracy* sebesar 80,22% *Precision* 84,87% dan *Recall* sebesar 90,74%.

Kata Kunci : Sentimen analisis, generasi z, gay, algoritma naïve bayes

1. PENDAHULUAN

Pada Era perkembangan Zaman tidak bisa di hindari bahwa semakin bertambahnya waktu dunia digital semakin berkembang dan maju, saat ini perkembangan teknologi informasi semakin canggih dan peminat pengguna yang semakin bertambah apa lagi saat ini generasi z sangat meminati dunia media sosial untuk bersosial, contohnya dengan pengguna twitter, namun semakin banyak kemudahan dan informasi yang tidak bisa dipilih oleh generasi z tidak menutup kemungkinan bahwa mereka mengenal tentang kehidupan gay, fenomena gay ini sangat berkembang dan lebih terbuka dengan adanya media sosial yang bebas digunakan.

Permasalahan saat ini bagaimana mengetahui pandangan dan sikap generasi z terhadap kehidupan gay pada twitter dan bagaimana mengetahui analisis sentimen sejauh mana pengetahuan kehidupan gay berupa tanggapan negatif, dan positif. Perkembangan kaum gay pada generasi z di Indonesia sedang mencerminkan salah satu pergeseran budaya Indonesia yang menarik perhatian Masyarakat. Kebebasan bicara kaum gay dan pengguna media sosial twitter memiliki berbagai pengaruh terhadap generasi z yang awalnya tidak mengetahui apa itu gay dan bisa mengetahuinya bahkan bisa saja mengikuti perilakunya, tujuan dari penelitian ini untuk mengetahui akibat dari pengetahuan kehidupan gay pada generasi z dan untuk

mengetahui sentimen opini negatife, dan positif. pada pengetahuan kehidupan gay pada generasi z.

Penelitian yang dilakukan oleh Anjani dan kawan-kawan dengan judul karya ilmiah Analisis Sentimen kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naïve bayes, tujuan dari penelitian jurnal tersebut untuk menggambarkan pendapat di internet mengenai positif dan negatif pada kaum LGBT pada penelitian jurnal tersebut mengambil data sebanyak 1579 data dari twitter dan menggunakan Algoritma Naïve Bayes untuk mengklarifikasi komentar Masyarakat pada twitter melalui klasifikasi data teks. [1]

Tujuan utama dari penelitian ini adalah Untuk menganalisis data terhadap pandangan generasi z terhadap kehidupan gay pada twitter dengan tanggapan positif dan negatif, serta untuk mengetahui Berapa nilai akurasi, presisi, dan recall pada analisis sentimen kehidupan gay berupa tanggapan negatife, dan positif, pada generasi z terhadap kehidupan gay menggunakan algoritma naïve bayes.

Untuk menganalisis data terhadap pandangan generasi z terhadap kehidupan gay pada twitter dengan tanggapan positif dan negatif, serta untuk mengetahui Berapa nilai akurasi, presisi, dan recall pada analisis sentimen kehidupan gay berupa tanggapan negatife, dan positif, pada generasi z terhadap kehidupan gay. penelitian ini menggunakan data teks dengan mendapatkan data dari twitter dengan cara melakukan

scraping data menggunakan tool google colab dan Microsoft excel sebanyak 1500 data. hasil penelitian ini menggunakan algoritma naïve bayes classifier dikarenakan cocok untuk penelitian analisis sentimen sebagai metode klasifikasi kedalam kategori positif dan negatif serta melakukan pengolahan data menggunakan rapidminer yang menghasilkan sentimen positif, dan negatif.

Penelitian oleh Hadiyan & Bachtiar yang berjudul Analisis Sentimen Aspek pada Opini Wisatawan menggunakan Metode Naïve Bayes (Studi Kasus: Perum Perhutani) membahas tentang pendapat wisatawan perum perhutani, perum perhutani salah satu tempat wisata namun tidak banyak dikenal dikarenakan belum ada pengelolaan yang baik pada tempat wisata tersebut dan belum dapat melakukan optimal karena kurangnya data dan informasi tentang perhutani.[2]

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh Khoirul Insan dkk, dengan judul Analisis sentimen aplikasi brimo pada ulasan pengguna google play menggunakan algoritma naïve bayes, pada penelitian ini membahas tentang opini pengguna aplikasi brimo berupa komentar, peneliti menggunakan data berupa data teks, mendapatkan opini dari pengguna aplikasi brimo dengan penelitian ini bertujuan untuk mencari analisis sentimen klasifikasi positif dan negatif dengan hasil penelitian tentang sentiment positif dan negatif menghasilkan proses akhir dengan hasil klasifikasi menggunakan metode algoritma naïve bayes dengan hasil didapatkan *accuracy* menggunakan algoritma naïve bayes mendapatkan tingkat *accuracy* dengan jumlah 84,52% *precision* 82,51% dan *recall* 87,62%. [3]

Penelitian yang berjudul Sentimen Analisis Terhadap Kaum Gay Pada Twitter Menggunakan Metode Naïve Bayes penelitian ini untuk menganalisis sentiment tweet pada kaum gay dengan menggunakan metode naïve bayes data yang diolah dengan melakukan preprocessing dan hasil yang di dapatkan dari 1635 data yang diambil, tweet yang bernilai negative berdasarkan nilai polaritas sebanyak 1505. [4].

Penelitian dari Edward Charitoun & Gawa dengan judul Penggunaan Media Sosial Sebagai Simbol Dalam Mendukung Hubungan LGBT, penelitian ini membahas mengenai penggunaan media sosial dalam mendukung hubungan kaum LGBT pada media sosial tiktok dalam penelitiannya memiliki masalah Semakin populer media komunikasi seperti tiktok semakin banyak yang menayangkan konten dewasa dan semakin bebas dalam berinteraksi penelitian ini menggunakan metode kualitatif dengan melakukan pendekatan studi literatur. [5]

Penelitian yang dilakukan oleh Sidabalok dkk yang berjudul Fenomena Komunikasi Kaum Gay Di Era Digital membahas tentang kaum gay yang

berkomunikasi lewat media sosial dengan menggunakan aplikasi Blued, dimana aplikasi ini digunakan oleh kaum gay untuk memuaskan Hasrat seksualnya dan bertemu dengan pasangan sesama jenisNyaKaum gay membutuhkan suatu media alternatif untuk berinteraksi dan komunikasi dengan tujuan untuk memenuhi kebutuhan sosial dan seksualnya, Penelitian ini menggunakan metode penelitian kualitatif deskriptif.[6]

2.2. Naïve Bayes

Naive Bayes sebuah metode yang tidak memiliki aturan dan menggunakan cabang matematika yang sering disebut dengan teori probabilitas untuk mendapatkan peluang setinggi mungkin dengan melihat kemunculan jumlah setiap klasifikasi dalam data yang digunakan. Dalam melakukan pengolahan pengembangan basis data, Algoritma Naive Bayes melibatkan pembelajaran yang diawasi dengan jenis pembelajaran mesin yang perlu membutuhkan sampel sebagai data pelatihan label. dikelompokkan menjadi dua bagian, yaitu. Klasifikasi dan regresi. Klasifikasi saat variabel menjadi kategori, panas dan dingin, sakit atau tidak sakit dan lain sebagainya.[3]

2.3. Media Sosial

Di era digital media sosial semakin banyak digunakan di Indonesia bahkan di dunia. Media sosial merupakan media berbasis online yang digunakan untuk berinteraksi secara individu maupun kelompok dengan efektif dan efisien yang bersifat tak terbatas. Media sosial sebagai tempat mengajak yang dapat mengubah pandangan dan perilaku publik. Setiap tahunnya, media sosial selalu berkembang pesat dan sudah familiar terutama di kalangan generasi Y dan Z [7].

2.4. Pengertian Gay

Gay merupakan istilah yang digunakan untuk menggambarkan seorang pria yang secara seksual tertarik kepada sesama pria dan menunjukkan kepada komunitas yang berkembang diantara orang-orang yang mempunyai orientasi seksual yang sama, istilah gay biasanya dikontraskan dengan straight, biseksual adalah istilah yang digunakan untuk menggambarkan orang-orang yang tertarik pada dua jenis kelamin sekaligus, jadi tipe ini tertarik pada laki-laki dan tertarik juga pada Perempuan.[8]

2.5. Generasi Z

Sebagai generasi terbaru yang lahir setelah generasi Y atau generasi *millenials*, generasi Z dipilih karena dipandang sebagai generasi penerus bangsa di masa depan. Dari rentang usia yang dihasilkan oleh dua penelitian sebelumnya, diketahui bahwa tidak semua generasi Z termasuk dalam usia produktif. Beberapa di antaranya ada yang sudah memasuki dunia kerja, tetapi tidak sedikit pula yang masih berstatus pelajar atau bahkan anak-anak. Generasi z adalah generasi yang akan mendatang yang sangat

mempengaruhi berbagai hal baik dalam dunia komunikasi, industri, perdagangan dan yang lainnya jika berdasarkan usia generasi z saat ini bisa dikatakan yang berusia 7-24 tahun.[9]

2.6. Twitter

Twitter adalah sebuah platform media sosial yang mempunyai layanan microblogging yang dimana mengizinkan penggunanya untuk mengirim pesan secara tepat waktu Pesan di Twitter terkenal dengan nama tweet. Twitter sendiri aplikasi yang sangat disukai dari semua jenis kalangan baik kalangan biasa maupun teratas. Twitter merupakan jejaring sosial dan situs berita online Dimana pengguna bisa berkomunikasi satu sama lain. [10]

2.7. Tanggapan positif Labeling

Berikut adalah contoh tanggapan positif dan negatif menurut Rintowati & Wahyudi

Tabel 1. Hasil label Positif

Data	Wujud tanggapan positif	Ciri penanda tanggapan	Akun Instagram
1.	Saya guru sdh 25 tahun dan baru kali ini senang sekali dengan mendikbud yang sekarang. Maju terus pak	Senang atau gembira	@laila.w

Tabel 2. Hasil Label Negatif

Data	Wujud tanggapan positif	Ciri penanda tanggapan	Akun Instagram
1.	Assalamu'alaikum pak nadiem, saya sangat sedih pak kok saya tidak menerima bantuan PIP sedang saya mempunyai KIP, capek saya pak mengadu sana sini gak direspon	Tidak senang/tidak gembira	@seceamush

3. METODE PENELITIAN

Metode penelitian yang digunakan yaitu melalui pendekatan kuantitatif, penelitian kuantitatif yaitu penelitian yang memakai data berupa angka sebagai alat untuk menganalisis keterangan tentang apa yang ingin diketahui, pendekatan kuantitatif merupakan metode pendekatan yang biasanya menekankan analisisnya pada data-data numerik atau angka yang digunakan untuk meneliti banyaknya sampel atau populasi yang akan diuji. Adapun rumus dari Teroma Bayes Sebagai berikut :[3]

$$P(H|X) = \frac{P(B|A) \times P(H)}{P(X)}$$

- X : Data yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P(H\X) : Probabilitas hipotesis H berdasarkan kondisi X (*Posterior Probabbility*)
- P(H) : Probabilitas hipotesis H (*Prior Probability*)
- P(X\H) : Probabilitas X berdasarkan pada Hipotesis H
- P(X) : Probabilitas X .[3]

Metode penelitian ini mengembangkan sentimen analisis pengetahuan kehidupan gay pada generasi z menggunakan metode Algoritma Naïve Bayes menggunakan metode Teknik Analisis Data menggunakan metode penelitian KDD atau disebut juga dengan Knowledge Discover in Database (KDD) dimana Teknik Analisis Data pada metode KDD ini merupakan cara untuk mengekstraksi data dalam mendapatkan informasi atau pola dari suatu data sebelumnya.[3] Tahapan dalam metode KDD meliputi data *Selection, Preprocessing, Transformation, Data Mining, dan Evaluation.*

3.1. Data Selection

Pada tahap ini merupakan tahapan KDD dimana dilakukan seleksi atribut dan pengumpulan data, pengumpulan data diambil langsung dari twitter dengan melakukan scraping data menggunakan kata kunci ciri gay kemudian data yang didapatkan sebanyak 1500 data dengan menggunakan 12 atribut seperti atribut *created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id_str, conversation_id_str, username, dan tweet_url.* Kemudian hanya digunakan satu atribut untuk mengolah data dengan atribut *text.*

3.2. Preprocessing

Tahap menghapus data dan perbaikan data. Pada saat melakukan pengumpulan data yang didapatkan pada saat melakukan pengambilan data atau scraping biasanya data yang didapatkan pada saat itu tidak terstruktur dan masih mengandung karakter maupun simbol yang tidak dibutuhkan. Tujuan adanya proses preprocessing adalah untuk menghilangkan noise, dan dalam tahap preprocessing terdapat beberapa Tahapan yang diperlukan, seperti tahap *cleaning, case folding, tokenize, dan filtering.* Bertujuan untuk membersihkan data *text* agar mendapatkan data yang lebih baik dan lebih mudah diolah.

3.3. Transformation

Proses untuk mengubah data teks menjadi bentuk yang bisa dilakukan pengolahan pada tahapan data mining. Dengan adanya tahap proses untuk mengubah data ini agar mempermudah proses olah data pada tahap pengolahan Algoritma yang akan digunakan.

3.4. Data Mining

Pada tahap data mining tahap ini adalah proses untuk pengolahan data yang digunakan pada algoritma yang konsisten dengan alur data mining, pada tahapan sebuah data mining penelitian ini menggunakan klasifikasi analisis sentimen pada data pengetahuan gay dan ciri gay terhadap generasi z menggunakan algoritma naïve bayes.[3]

3.5. Evaluation

Tahap *evaluation* digunakan untuk mengevaluasi algoritma yang digunakan pada penelitian ini, Pada tahap *evaluation* pada penelitian ini akan menggunakan *confusion matriks* untuk mengetahui sebuah hasil dari performa akurasi yang dihasilkan oleh algoritma naïve bayes dengan hasil berupa *accuracy*, *precicion*, dan *recall*. [3] Dimana *confusion matrix* adalah salah satu metode untuk mengukur kinerja dari model klasifikasi.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian yang telah dilakukan pada penelitian ini adalah untuk memberikan tahapan proses bagaimana mengklasifikasikan suatu data teks mengenai opini atau pendapat pengguna media sosial twitter tentang ciri-ciri gay berupa tanggapan mereka yang mendukung tentang gay ataupun yang menolak tentang gay pada generasi z dengan menggunakan algoritma naïve bayes. dalam penelitian ini menggunakan proses pengujian pada perangkat lunak seperti Rappid Minner Studio 10.3.0-win64.

4.1. Data Selection

Data yang didapatkan adalah data pengguna twitter yang digunakan oleh generasi z, dari data yang

4.2. Cleansing

created_at	id_str	full_text
Mon Nov 13 00:14:46 +0000 2023	1,72E+18	@roundhatboy 1. orientasi seksual tu gk bsa ditularin, w gk punya temen gay tp perasaan ni tumbuh alami seiring waktu, 2. STD tu gk ciri khas kami, yg Cis-hetero jg sama2 beresiko 3. gmn kami bisa menuntut pemenuhan hak kami DI NEGARA DEMOKRASI klo gk dibolehin demonstrasi dgn aman?
Wed Nov 08 00:16:24 +0000 2023	1,72E+18	ih kamu suka Nagi Reo GK" Kepercayaan diri kurang, lemah, gk bisa narik peminat "Heh kontol ngelunjak banget!! Udh2 si Reo tuh gay, dia sukanya sama Nagi, gasuka cewek tolol" Berani, ciri orang sukses, dewasa"
Mon Nov 06 16:36:29 +0000 2023	1,72E+18	@akif_ax @f4kr0ls00d KAU yang dah kantoi kena exposed. Ngaku je lah kau gay and liberal, tapi masih ada ciri ciri islamik dalam diri & akan terus menyokong kepemimpinan PNPAS hingga ke jannah. Jangan pulak kau ingat nanti kau dapat 80k gay servants & 72 gay wives kat sorga nanti. Kau ngaku je, Nyah.
Mon Nov 06 05:16:36 +0000 2023	1,72E+18	Anj wkwk kaca mata, hobi mirror selfie, maen twitter, tembok kamar putih abu, ciri khas gay. Tapi gua ga gay bajing δŸ—ιδŸ—ι

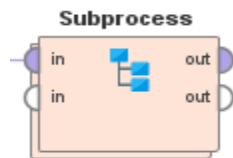
Proses pembersihan atau penghapusan operator tertentu untuk menghapus symbol, karakter, angka, dan kalimat yang tidak ingin digunakan, operator yang digunakan untuk penghapusan simbol digunakan operator *replace url*, *replace* pertama dinamakan *replace* satu digunakan untuk menghapus simbol-simbol berikut adalah gambar proses cleansing yaitu *subprocess* pada proses *subprocess* didalamnya ada

terkumpulkan penulis menggunakan penelitian metode kuantitatif. Dengan mengambil data secara langsung dengan scrapping data menggunakan kata kunci ciri gay. Mengunjungi situs resmi twitter dengan keyword ciri gay lalu kemudian dilakukan scraping dengan target link pada salah satu akun yang memiliki topik gay dengan mengambil data sebanyak 600 data dari tahun 2019. Setelah itu didapatkan data scraping dengan format csv menggunakan reverensi pengambilan data dari youtube dengan situs <https://helimisatria.com/blog/cara-mendapatkan-data-twitter-yang-besar>. Hasil penelusuran data tersebut dalam bentuk csv dan mendapatkan data menggunakan 12 atribut seperti atribut *created_at*, *id_str*, *full_text*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, dan *tweet_url*. Kemudian pada langkah ini dilakukan pemilihan atribut yang akan digunakan, sebelumnya dalam tahap pengambilan data melalui scrapping data yang didapatkan 1500 data dengan menggunakan 12 atribut. Seperti atribut *created_at*, *id_str*, *full_text*, *quote_count*, *reply_count*, *retweet_count*, *favorite_count*, *lang*, *user_id_str*, *conversation_id_str*, *username*, dan *tweet_url*. Selanjutnya melakukan penghapusan atribut yang digunakan hanya Text. berikut adalah Atribut Text pada Tabel 3.

Tabel 3. Atribut Text

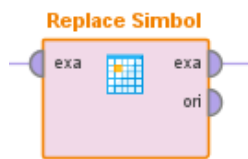
No	Text
1.	@budakfasyen sumpah perlu untuk gay yang ga segan ngelakuin pelecehan
2.	@kegblgnunfaedh Biarin, udah makan sendiri ga bagi2, mukanya menyerupai kaum Gay

proses *replace url1*, *replace url2*, *replace mention*, *replace simbol*, dan *trim*. Berikut adalah proses dari *subprocess* pada Gambar 1.



Gambar 1. Proses subprocess

Berikut adalah tahapan dari proses cleansing yaitu tahapan proses Replace Simbol dengan filter type memilih *single*, atribut *Text*, dan Replace What diisi dengan [~!@#%\$%^&*():;...-] bertujuan untuk menghapus Teks yang masih mempunyai simbol. Berikut adalah Gambar 2. proses Replace Simbol.



Gambar 2. Proses Replace Simbol

Untuk mengetahui perbedaan sebelum dan sesudah menggunakan operator replace simbol berikut adalah contoh perbedaannya pada Tabel 4.

Tabel 4. Replace simbol

Input proses	Output proses
DONE MAS BRO! Thanks mas bro outcallnya!! Open pijat Kota makassar OUTCALL(panggilan) Wa. 0895329438172	DONE MAS BRO Thanks mas bro outcallnya Open pijat Kota makassar OUTCALL panggilan Wa. 0895329438172
apagaaaa :~	apagaaaa

4.3. Labeling

Labeling adalah proses Setelah semua data telah dibersihkan, pelabelan yang dilakukan secara manual dengan membaca satu persatu kata atau kalimat yang memiliki arti positif atau negatif pada topik yang sedang dibahas seperti topik pada penelitian ini adalah topik tentang ciri gay. Data yang diolah sebanyak 1500 data namun setelah dilakukan pelabelan data dan proses cleansing sampel yang digunakan sebanyak 1010 data. Berikut adalah contoh hasil sebelum dan sesudah

Tabel 5. Hasil Input dan Output

Input proses	Output proses
Kalau aing open BO laku ga ya	kalau aing open bo laku ga ya
Boti Trending	boti trending

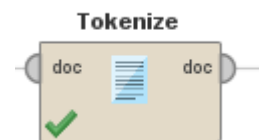
Tabel 6. Hasil input Output

Input proses	Output proses
cowo gw gay magnet kali ya	Negatif
Terus yg udh gay di dalem situ gmn wkwk diusir gitu? Ini apaan sih gk jelas apanya yg dilarang wkwk bikin peraturan kok menyudutkan komunitas	Positif

Pelabelan dikatan positif adalah mereka yang mendukung dengan adanya gay sedangkan mereka yang berpendapat negatif adalah mereka yang tidak setuju tentang gay pelabelan terdapat pada Tabel 6.

4.4. Tokenize

Tahap selanjutnya yang dilakukan adalah melakukan proses tokenisasi, Dimana proses pecahan dokumen menjadi kata berikut angka 0 adalah angka yang menunjukkan seberapa sering kata yang diulang, berikut adalah contoh sebelum dan sesudah proses tokenize dapat dilihat pada Gambar 3. Proses *Tokenize*.



Gambar 3. Proses Tokenize

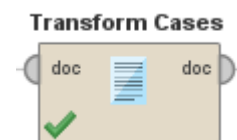
Berikut adalah contoh perbedaan sebelum dan sesudah melakukan proses *Tokenize* berikut adalah Tabel 7. Tahapan *Tokenize*.

Tabel 7. Tahap Tokenize

Input proses		Output proses		
Text	Senti men	Text	Senti men	ABANG NYA
obat untuk yang gay	Positif	obat untuk yang gay	Positif	0
Astaghfir ullah gay	Negatif	Astaghfir ullah gay	Negatif	0

4.5. Transform cases

Pada tahap *transform case* atau case folding yaitu proses mengubah teks menjadi huruf Akecil dari dokumen semua yang awalnya besar akan menjadi kecil dengan proses transform case berikut adalah contoh transform case pada gambar 4. Proses *Transform Cases*.



Gambar 4. Tahap Transform case

Berikut adalah contoh tahapan sebelum dan sesudah melakukan proses *Transform Cases* pada Tabel 4.8. *Tahap Transform Cases*.

4.6. Stopword Removal

Adalah proses Dimana melakukan proses pengapusan data atau kalimat yang tidak penting atau tidak memiliki makna namun tidak berpengaruh pada suatu kalimat contohnya "ada" "yang" "aku" berikut adalah contoh bentuk kalimat *stopword removal* dengan melakukan proses dari data csv Bahasa Indonesia. Berikut adalah proses sebelum dan sesudah

melakukan proses *Stopword Removal* terdapat pada Tabel 8.

Tabel 8. Tahap Stopword Removal

Input proses	Output proses
gak semua orang gay itu cabul sama halnya dengan tidak semua orang islam itu teroris	gak orang gay cabul halnya orang islam teroris
obat untuk yang gay	obat gay

4.7. Filter Tokens (by length)

Pada langkah ini dilakukan tahap penghapusan kata atau sejumlah kata singkatan yang minimal karakternya sudah ditentukan sebelum digunakan, pada penelitian ini menggunakan Panjang minimal 4 karakter dan maksimal 25 karakter. Berikut adalah gambar 4.10. Proses *Filter Tokens (by length)*. Berikut adalah contoh dari proses filter token *By Length* Sebelum dan Sesudah menggunakan *Filter Tokens (By Length)* pada Tabel 9.

Tabel 9. Tahapan Filter Tokens (by length)

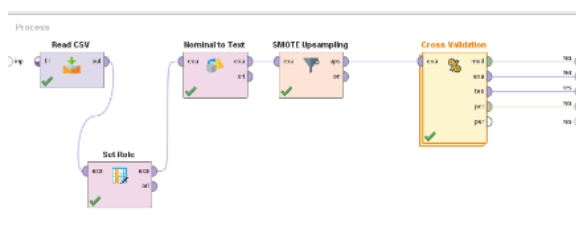
Input proses	Output proses
obat untuk yang gay	obat
punya ex pasangan gay	pasangan
hans yang menerima dia gay	hans menerima

4.8. Transformation

Tahap mengubah data menjadi bentuk yang dapat diolah pada tahapan data mining adalah mengubah data nominal menjadi sebuah *text*.

4.9. Data Mining

Pada tahap ini dilakukan tahapan model algoritma yang akan digunakan, pada penelitian ini menggunakan model naïve bayes diantaranya proses membaca excel yang berfungsi menampung data train lalu dihubungkan pada *set role* yang berguna untuk merubah dara role menjadi tabel kemudian dihubungkan dengan proses *document from data* yang didalamnya terdapat sub proses *Tokenize*, *Transform cases*, kemudian *cross validation* untuk menguji model naïve bayes dengan pengujian 1-10 kali. Dapat dilihat Pada Gambar 5.



Gambar 5. Proses Model Naive Bayes

Berdasarkan hasil dari klasifikasi Naïve Bayes Setelah melakukan perhitungan manual hasil persentasi sentimen dengan jumlah ulasan positif lebih banyak positif pada pembahasan disini adalah mereka yang setuju atau mendukung tentang gay dari pada jumlah ulasan negatif, jumlah ulasan sentimen positif

sebesar 75,74%, dan jumlah ulasan sentimen negatif sebesar 24,25%, pendapat negatif disini adalah mereka yang tidak setuju tentang gay.

Teks dari total data teks sentimen yang didapatkan melalui scraping yang awalnya berjumlah 1010 data setelah melalui proses tokenize dan labeling didapatkan data 1010 data teks. Berikut adalah Gambar 6. Diagram Sentimen.



Gambar 6. Diagram Sentimen

Berikut presentasi sentimen positif dan negatif dari hasil klasifikasi dengan hasil sentimen positif dibagi seluruh dokumen dan sentimen negatif dibagi seluruh dokumen. Berikut persentasi sentimen negatif dan positif dibawah ini.

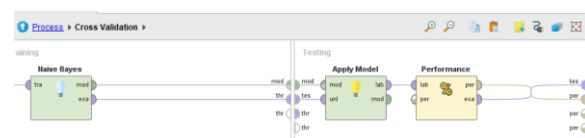
$$\frac{765}{1010} \times 100\% = 75,74\%$$

$$\frac{245}{1010} \times 100\% = 24,25\%$$

Setelah melakukan perhiutngan manual hasil persentasi sentimen dengan jumlah ulasan positif lebih banyak dari pada jumlah ulasan negatif, jumlah ulasan sentimen positif sebesar 75,74%, dan jumlah ulasan sentimen negatif sebesar 24,25%. artinya dari hasil keseluruhan klasifikasi terdapat hasil prediksi sentimen positif lebih besar daripada sentimen negatif tentang pengetahuan kehidupan gay.

4.10. Evaluation

Tahapan ini dilakukan untuk menghitung evaluasi dari model yang akan dibuat, dalam penelitian ini menggunakan model naïve bayes dengan cross validation yang didalamnya dihubungkan dengan *apply model* dan *performance* untuk dihitung terhadap hasil dari naïve bayes. Berikut adalah gambar 7. Proses Validation Naïve Bayes.



Gambar 7. Proses Validation Naive Bayes

Berdasarkan hasil evaluasi model naïve bayes pada penelitian ini mendapatkan hasil bahwa *accuracy* sebesar 80,22% bisa dilihat pada Gambar 8. Hasil *Accuracy* Naïve bayes.

	true positif	true negatif	class precision
pred positif	533	71	88.25%
pred negatif	232	694	74.95%
class recall	69.67%	90.72%	

Gambar 8. Hasil Accuracy Naive Bayes

Berikut adalah gambar hasil dari precision pada Gambar 9.

	true positif	true negatif	class precision
pred positif	533	71	88.25%
pred negatif	232	694	74.95%
class recall	69.67%	90.72%	

Gambar 9. Hasil Precision

Selanjutnya adalah nilai Recall berikut adalah Gambar Nilai Recall yang didapatkan 90,74% terdapat pada Gambar 10.

	true positif	true negatif	class precision
pred positif	533	71	88.25%
pred negatif	232	694	74.95%
class recall	69.67%	90.72%	

Gambar 10. Hasil Recall

Didapatkan hasil pengujian klasifikasi dengan jumlah data sentimen 1010 data menggunakan metode Algoritma naïve bayes didapatkan accuracy sebesar 80,22% Precision 84,87% dan Recall sebesar 90,74%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian data teks analisis sentimen generasi z terhadap pengetahuan gay yang menggunakan algoritma naïve bayes menggunakan Discovery in Database (KDD) Keakuratannya diuji menggunakan 1010 data dari pengujian yang didapatkan generasi z yang mengetahui tentang ciri gay dan apa itu gay pada twitter adalah mereka yang setuju atau mendukung dengan pendapat tentang gay dan ciri gay pada twitter, namun tidak sedikit juga generasi z yang memberikan pendapatnya tidak setuju dengan ciri-ciri gay yang ada pada twitter Dan hasil pengujian klasifikasi menggunakan metode Algoritma naïve bayes didapatkan accuracy sebesar 80,22% Precision 84,87% dan Recall sebesar 90,74%. Saran untuk penelitian yang akan mendatang gunakanlah Algoritma yang berbeda.

DAFTAR PUSTAKA

[1] A. Anjani, A. Chamid, and A. Murti, “Analisis Sentimen Kaum LGBT pada Media Sosial

Twitter Menggunakan Algoritma Naïve Bayes,” *JTINFO J. Tek. Inform.*, vol. 1, no. 2, pp. 1–8, 2022.

[2] F. G. Hadiyan and F. A. Bachtiar, “Analisis Sentimen Aspek pada Opini Wisatawan menggunakan Metode Naïve Bayes (Studi Kasus: Perum Perhutani),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 9, pp. 4007–4014, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>

[3] M. K. Khoiril Insan, U. Hayati, and O. Nurdiawan, “Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.

[4] A. Pramitha, E. Supriyati, and T. Listyorini, “Sentimen Analisis Terhadap Kaum Gay Pada Twitter Menggunakan Metode Naive Bayes,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 1, pp. 1–4, 2021, doi: 10.26740/jieet.v5n1.p1-4.

[5] E. Charitoun and S. Gawa, “Penggunaan Media Sosial Sebagai Simbol dalam Mendukung Hubungan,” vol. 05, no. 04, pp. 15598–15608, 2023.

[6] R. Sidabalok, S. Telussa, P. Fakultas, I. Sosial, and I. Komunikasi, “Fenomena Komunikasi Kaum Gay Di Era Digital,” vol. 01, no. 02, pp. 1–18, 2022.

[7] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.

[8] F. S. Dhamayanti, “Pro-Kontra Terhadap Pandangan Mengenai LGBT Berdasarkan Perspektif HAM, Agama, dan Hukum di Indonesia,” *Ikat. Penulis Mhs. Huk. Indones. Law J.*, vol. 2, no. 2, pp. 210–231, 2022, doi: 10.15294/ipmhi.v2i2.53740.

[9] A. C. Amalia and G. Sagita, “Analisa Pengaruh Influencer Social Media Terhadap Keputusan Pembelian Konsumen Generasi Z di Kota Surabaya,” *J. Sos. J. Penelit. Ilmu-Ilmu Sos.*, vol. 20, no. 2, pp. 51–59, 2019, doi: 10.33319/sos.v20i2.42.

[10] D. A. Wulandari, R. Rohmat Saedudin, and R. Andreswari, “Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat pada RUU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes,” *J. e-Proceeding Eng.*, vol. 8, no. 5, pp. 9007–9016, 2021.