JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.10387

ANALISIS SENTIMEN OPINI PUBLIK TERHADAP STAND-UP COMEDY "MENS REA" DI MEDIA SOSIAL X DENGAN NAIVE BAYES

Rizki Juli Aditia[1*], Siska[2], Billy Ibrahim Hasbi [3]

[1,2,3] Universitas Singaperbangsa Karawang, Jl. HS. Ronggo Waluyo, Pseurjaya, Telukjambe Timur, Karawang, Jawa Barat, 41361.

Corespondent Email:
kiritokunoflegendry@gmail.com

Abstrak: *Stand-up comedy* Mens Rea karya Pandji Pragiwaksono memunculkan beragam tanggapan masyarakat di media sosial X yang mencerminkan sentimen positif, negatif, dan netral. Penelitian ini bertujuan menganalisis sentimen opini publik terhadap pertunjukan tersebut menggunakan algoritma Multinomial Naive Bayes. Data penelitian diperoleh melalui proses *crawling* pada media sosial X sebanyak 3.019 tweet. Tahapan penelitian meliputi *data selection*, pelabelan sentimen menggunakan pendekatan *lexicon-based* yang divalidasi oleh validator ahli, *preprocessing* (*cleansing, case folding, tokenizing, stopword removal,* dan *stemming*), pembobotan kata menggunakan TF-IDF, serta klasifikasi menggunakan Multinomial Naive Bayes. Evaluasi dilakukan menggunakan Confusion Matrix pada lima skenario pembagian data *training* dan *testing* (90:10, 80:20, 70:30, 60:40, dan 50:50). Hasil penelitian menunjukkan bahwa performa terbaik diperoleh pada skenario 80:20 dengan nilai accuracy sebesar 85,79%, sedangkan nilai F1-score tertinggi sebesar 0,9231 diperoleh pada kelas sentimen netral. Secara keseluruhan, kombinasi TF-IDF dan Multinomial Naive Bayes mampu memberikan performa klasifikasi yang baik dan stabil dalam menganalisis sentimen opini publik terhadap *stand-up comedy* Mens Rea di media sosial X.

Abstract: The Mens Rea stand-up comedy performance by Pandji Pragiwaksono has generated diverse public responses on the X social media platform, reflecting positive, negative, and neutral sentiments. This study aims to analyze public sentiment toward the performance using the Multinomial Naive Bayes algorithm. Research data were collected through a crawling process on X, resulting in 3,019 tweets. The research stages included data selection, sentiment labeling using a lexicon-based approach validated by an expert, preprocessing (cleansing, case folding, tokenizing, stopword removal, and stemming), TF-IDF term weighting, and sentiment classification using Multinomial Naive Bayes. Model performance was evaluated using a Confusion Matrix with five train-test split scenarios (90:10, 80:20, 70:30, 60:40, and 50:50). The best performance was achieved using the 80:20 split, with an accuracy of 85.79%, while the highest F1-score of 0.9231 was obtained for the neutral sentiment class. Overall, the combination of TF-IDF and Multinomial Naive Bayes demonstrated stable and effective performance for public sentiment classification of the Mens Rea stand-up comedy on the X platform.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat berinteraksi, berkomunikasi, serta menyampaikan pandangan di ruang publik. Kehadiran media sosial memungkinkan pengguna tidak hanya menerima informasi, tetapi juga memproduksi dan menyebarkan informasi secara real-time, sehingga menjadi ruang publik digital yang merepresentasikan dinamika opini masyarakat terhadap berbagai isu [1]. Karakteristik media sosial yang mampu merekam respons pengguna secara spontan menghasilkan data teks dalam jumlah besar yang mencerminkan sikap, penilaian, maupun emosi publik. Namun, volume data yang terus meningkat serta sifatnya yang tidak terstruktur menyebabkan analisis opini secara manual menjadi kurang efektif dan berpotensi menghasilkan penilaian yang tidak konsisten.

Salah satu media sosial yang banyak dimanfaatkan sebagai sumber data opini publik adalah Platform X. Sebagai platform microblogging, X memungkinkan pengguna menyampaikan pandangan melalui unggahan teks secara langsung sehingga respons terhadap suatu isu dapat tersebar dengan cepat dan mencerminkan persepsi aktual masyarakat [2]. Selain digunakan untuk membahas isu politik maupun kebijakan publik, Platform X juga menjadi ruang diskusi berbagai fenomena budaya populer, termasuk pertunjukan stand-up comedy. Salah satu pertunjukan yang menarik perhatian publik adalah Mens Rea karya Pandji Pragiwaksono yang mengangkat tema hukum dan kritik sosial. Perbedaan interpretasi masyarakat terhadap materi yang disampaikan memunculkan beragam opini di Platform X, terlebih setelah muncul pemberitaan mengenai respons Pandji terhadap laporan hukum yang berkaitan dengan materi tersebut. Fenomena ini menunjukkan bahwa Mens Rea tidak hanya dipandang sebagai hiburan, tetapi juga memunculkan diskusi mengenai kebebasan berekspresi di ruang publik digital.

Beragamnya opini yang muncul menghasilkan kecenderungan sentimen positif, negatif, maupun netral. Karakteristik data yang berbentuk teks tidak terstruktur, menggunakan bahasa informal, dan mengandung subjektivitas yang tinggi menjadikan analisis sentimen sebagai pendekatan yang relevan untuk mengidentifikasi kecenderungan opini masyarakat [3]. Berbagai penelitian terdahulu menunjukkan bahwa teknik text mining dengan algoritma Naive Bayes mampu mengklasifikasikan sentimen pada data Platform X secara efektif [4], sedangkan penggunaan pembobotan TF-IDF terbukti mampu meningkatkan representasi fitur sehingga mendukung kinerja klasifikasi [5]. Selain itu, Platform X juga dinilai sebagai sumber data yang efektif untuk opinion mining karena mampu merepresentasikan opini publik terhadap berbagai fenomena [6]. Meskipun demikian, sebagian besar penelitian masih berfokus pada isu politik, pelayanan publik, maupun produk, sedangkan penelitian mengenai sentimen masyarakat terhadap konten budaya populer, khususnya pertunjukan stand-up comedy Mens Rea, masih relatif terbatas.

Berdasarkan kondisi tersebut, penelitian ini mengusulkan analisis sentimen opini publik terhadap pertunjukan stand-up comedy Mens Rea di Platform X menggunakan algoritma Naive Bayes dalam kerangka Knowledge Discovery in Databases (KDD). Proses penelitian meliputi seleksi data, preprocessing (case folding, tokenizing, normalization, stopword removal, dan stemming), pembobotan TF-IDF, klasifikasi menggunakan Naive Bayes, serta evaluasi model menggunakan accuracy, precision, recall, dan F1-score [7]. Penelitian ini diharapkan mampu memberikan gambaran yang lebih objektif mengenai kecenderungan sentimen masyarakat terhadap konten stand-up comedy yang mengandung kritik sosial sekaligus memperluas penerapan analisis sentimen pada fenomena budaya populer di media sosial.

2. TINJAUAN PUSTAKA

Analisis sentimen merupakan salah satu teknik dalam *text mining* yang digunakan untuk mengklasifikasikan opini pada data teks ke dalam kategori positif, negatif, maupun netral. Teknik ini memanfaatkan Natural Language Processing (NLP) dan algoritma machine learning untuk memperoleh informasi dari data tidak terstruktur sehingga mampu menggambarkan kecenderungan opini masyarakat terhadap suatu fenomena [8]. Pada penelitian ini, analisis sentimen digunakan untuk mengidentifikasi opini publik terhadap pertunjukan stand-up comedy Mens Rea

berdasarkan data yang diperoleh dari Platform X.

Platform X merupakan media sosial berbasis microblogging yang memungkinkan pengguna menyampaikan opini melalui unggahan teks secara real-time. Karakteristik tersebut menjadikan Platform X sebagai salah satu sumber data yang banyak digunakan dalam penelitian analisis sentimen karena mampu merepresentasikan respons masyarakat terhadap berbagai isu yang berkembang [9]. Data yang diperoleh kemudian diolah menggunakan pembobotan TF-IDF dan diklasifikasikan menggunakan algoritma Naive Bayes.

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang mengacu pada Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen. Algoritma ini banyak digunakan dalam klasifikasi data teks karena memiliki proses komputasi yang sederhana, efisien, dan mampu memberikan performa yang baik pada berbagai penelitian analisis sentimen [10]. Probabilitas posterior pada algoritma Naive Bayes dihitung menggunakan Persamaan (1).

$$P(X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Keterangan:

$P(H \mid X)$ = probabilitas posterior.

$P(H)$ = probabilitas awal (*prior*) suatu kelas.

$P(X \mid H)$ =probabilitas kemunculan data terhadap kelas.

$P(X)$ =probabilitas keseluruhan data.

Tahap evaluasi dilakukan menggunakan Confusion Matrix untuk mengukur performa model klasifikasi. Confusion Matrix menghasilkan empat metrik evaluasi, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. [11], setiap metrik memberikan informasi yang berbeda mengenai kemampuan model, sehingga penggunaannya secara bersamaan mampu memberikan gambaran kinerja klasifikasi yang lebih komprehensif. Persamaan yang digunakan untuk menghitung keempat metrik tersebut ditunjukkan pada Persamaan (2)–(5).

1. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ (2)

2. $\text{Precision} = \frac{TP}{TP+FP}$ (3)

3. $\text{Recall} = \frac{TP}{TP+FN}$ (4)

4. $\text{F1-Score} = \frac{2 \times precision \times Recall}{Precision+Recall}$ (5)

Keterangan:

TP (*True Positive*) = jumlah data yang diprediksi benar pada kelas positif.

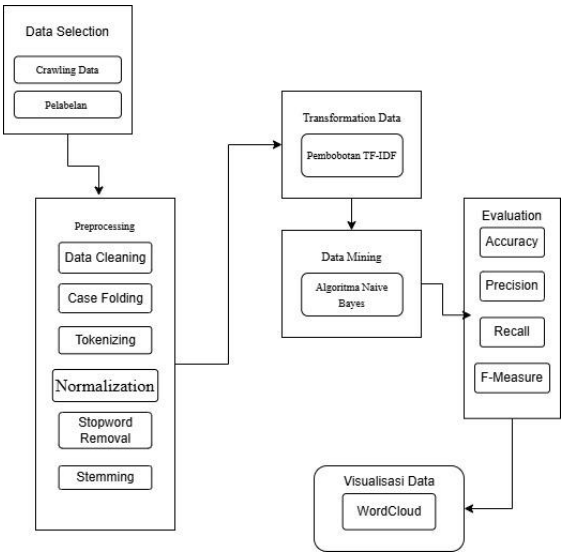
TN (*True Negative*) = jumlah data yang diprediksi benar pada kelas negatif.

FP (*False Positive*) = jumlah data yang diprediksi positif tetapi sebenarnya bukan kelas tersebut.

FN (*False Negative*) = jumlah data yang diprediksi bukan kelas tersebut tetapi sebenarnya termasuk dalam kelas tersebut.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) sebagai kerangka kerja dalam proses analisis sentimen opini publik terhadap pertunjukan stand-up comedy Mens Rea. Pendekatan KDD dipilih karena telah menyediakan tahapan pengolahan data yang sistematis, dimulai dari pengumpulan data hingga evaluasi dari hasil klasifikasi. Tahapan penelitian meliputi Data Selection, Preprocessing, Transformation, Data Mining, Evaluation, dan Visualisasi Data. Alur penelitian yang digunakan pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

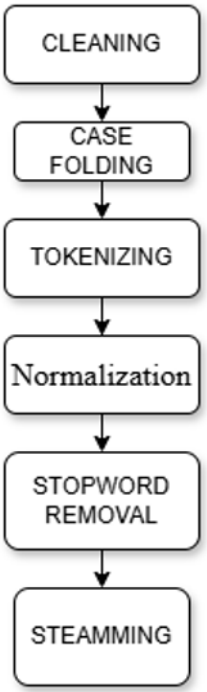
3.1 *Data Selection*

Tahap Data Selection dilakukan untuk memperoleh data yang akan digunakan pada proses analisis sentimen. Pengumpulan data dilakukan melalui proses crawling pada Platform X menggunakan library tweet-harvest dengan kata kunci "Stand Up Comedy Mens Rea". Proses pengambilan data dilakukan pada periode 27 Desember 2025 hingga 27 Februari 2026, sehingga diperoleh sebanyak 3.020 tweet sebagai dataset penelitian.

Data hasil crawling selanjutnya diseleksi untuk menghilangkan data duplikat, unggahan yang tidak relevan, serta data yang tidak sesuai dengan ruang lingkup penelitian. Setelah proses seleksi selesai, setiap tweet diberi label sentimen menggunakan metode Lexicon Based sehingga terbentuk tiga kategori sentimen, yaitu positif, negatif, dan netral. Data yang telah melalui proses pelabelan kemudian digunakan sebagai dataset pada tahapan preprocessing dan proses klasifikasi menggunakan algoritma Naive Bayes.

3.2 *Preprocessing*

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum proses klasifikasi. Pada tahap ini dilakukan beberapa proses, yaitu Data Cleaning**,** Case Folding**,** Tokenizing**,**Normalization**,** Stopword Removal**,** dan Stemming. Alur tahapan *preprocessing* pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Tahapan Preprocessing

Tahapan *preprocessing* yang diterapkan pada penelitian ini meliputi:

A. Cleaning :Data *cleaning* dilakukan untuk menghapus URL, *mention*, *hashtag*, emoji, tanda baca, angka, dan karakter lain yang tidak memiliki pengaruh terhadap proses analisis sentimen.
B. Case Folding :Tahap *case folding* bertujuan mengubah seluruh huruf menjadi huruf kecil (*lowercase*) sehingga penulisan kata menjadi seragam.
C. Tokenizing :*Tokenizing* dilakukan dengan memecah kalimat menjadi kumpulan kata (*token*) sehingga memudahkan proses pengolahan data pada tahap berikutnya.
D. Normalization : Tahap *Normalization* digunakan untuk mengubah kata tidak baku, singkatan, atau bahasa gaul menjadi kata baku sesuai kamus normalisasi.
E. Stopword Removal : Tahap *stopword removal* bertujuan menghapus kata-kata umum yang tidak memiliki makna penting dalam proses klasifikasi sentimen.
F. Stemming : *Stemming* dilakukan untuk mengubah setiap kata menjadi bentuk

kata dasar sehingga variasi kata dapat direpresentasikan dalam bentuk yang seragam.

3.3 *Transformasi Data*

Tahap Data Transformation dilakukan untuk mengubah data teks hasil preprocessing menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Pada penelitian ini digunakan metode Term *Frequency–Inverse Document Frequency (TF-IDF)* untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan distribusinya pada seluruh kumpulan dokumen. Hasil pembobotan TF-IDF kemudian digunakan sebagai fitur masukan (input feature) pada proses klasifikasi menggunakan algoritma Naive Bayes [12].

Selanjutnya, dataset yang telah dibagi menggunakan metode *train-test split* ke dalam lima skenario pembagian data untuk mengevaluasi pengaruh proporsi data pelatihan dan data pengujian terhadap kinerja model. Adapun skenario pembagian data yang digunakan adalah sebagai berikut.

- 90% data training dan 10% data testing
- 80% data training dan 20% data testing
- 70% data training dan 30% data testing
- 60% data training dan 40% data testing
- 50% data training dan 50% data testing

3.4 *Data Mining*

Tahap Data Mining dilakukan menggunakan algoritma Naive Bayes untuk membangun model klasifikasi berdasarkan data latih pada setiap skenario pembagian data. Model yang telah dibangun kemudian digunakan untuk memprediksi kelas sentimen pada data uji sehingga menghasilkan klasifikasi ke dalam kategori positif, negatif, dan netral. Algoritma Naive Bayes digunakan karena memiliki proses komputasi yang efisien dan telah terbukti memberikan performa yang baik pada berbagai penelitian klasifikasi sentimen berbasis teks [13].

3.5 *Evaluation*

Tahap Evaluation bertujuan untuk mengukur performa model klasifikasi yang dihasilkan pada setiap skenario pembagian data. Proses evaluasi dilakukan menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score sebagai indikator dalam menilai kemampuan model mengklasifikasikan sentimen secara tepat. Penggunaan keempat metrik tersebut memberikan evaluasi yang lebih menyeluruh terhadap hasil klasifikasi, sehingga dapat digunakan untuk membandingkan performa model pada setiap skenario pembagian data [14].

3.6 *Visualisai Data*

Tahap Visualisasi Data dilakukan menggunakan WordCloud untuk menampilkan kata-kata yang paling dominan berdasarkan frekuensi kemunculannya pada data sentimen. Visualisasi ini bertujuan memberikan gambaran awal mengenai pola kemunculan kata serta membantu mengidentifikasi istilah yang paling sering digunakan dalam opini publik terhadap pertunjukan *stand-up comedy* Mens Rea. Penggunaan WordCloud dinilai mampu menyajikan hasil analisis teks secara visual sehingga memudahkan proses interpretasi terhadap kecenderungan sentimen yang terbentuk [15].

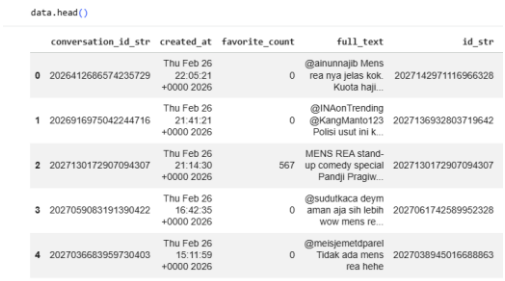
4. HASIL DAN PEMBAHASAN

Pada bab ini disajikan hasil penerapan metode Knowledge Discovery in Databases (KDD) pada analisis sentimen opini publik terhadap pertunjukan *stand-up comedy* Mens Rea di Platform X. Tahapan penelitian meliputi pengumpulan data, pada pelabelan sentimen menggunakan Lexicon Bassed, proses *preprocessing*, pembobotan kata TF-IDF, dan klasifikasi menggunakan algoritma Naive Bayes, serta evaluasi model pada lima skenario pembagian data. Hasil pada setiap tahapan dibahas untuk menunjukkan performa model dalam mengklasifikasikan sentimen.

4.1. Data Selection

1. Crawling

Pengumpulan data dilakukan melalui proses *crawling* pada Platform X menggunakan kata kunci *"Stand Up Comedy Mens Rea"* selama periode 27 Desember 2025 hingga 27 Februari 2026. Berdasarkan proses tersebut diperoleh sebanyak 3.020 tweet yang selanjutnya digunakan sebagai dataset pada penelitian ini. Contoh data hasil *crawling* ditunjukkan pada Gambar 3.



data.head()

	conversation_id_str	created_at	favorite_count	full_text	id_str
0	2026412686574235729	Thu Feb 26 22:05:21 +0000 2026	0	@ainunnajib Mens rea nya jelas kok. Kuota haji...	2027142971116966328
1	2026916975042244716	Thu Feb 26 21:41:21 +0000 2026	0	@INAonTrending @KangManto123 Polisi usut ini k...	2027136932803719642
2	2027130172907094307	Thu Feb 26 21:14:30 +0000 2026	567	MENS REA stand-up comedy special Pandji Pragiw...	2027130172907094307
3	2027059083191390422	Thu Feb 26 16:42:35 +0000 2026	0	@sudutkaca deym aman aja sih lebih wow mens re...	2027061742589952328
4	2027036683959730403	Thu Feb 26 15:11:59 +0000 2026	0	@meisjemetdparel Tidak ada mens rea hehe	2027038945016688863

Gambar 3. Hasil crawling

2. Pelabelan

Setelah proses seleksi data, setiap tweet diberi label sentimen menggunakan metode Lexicon Based berdasarkan nilai polaritas kata yang terdapat pada teks. Proses pelabelan menghasilkan tiga kategori sentimen, yaitu positif, negatif, dan netral. Berdasarkan hasil pelabelan terhadap 3.020 tweet, diperoleh 2.453 tweet dengan sentimen netral, 211 tweet dengan sentimen positif, dan 186 tweet dengan sentimen negatif. Hasil tersebut menunjukkan bahwa mayoritas opini pengguna terhadap pertunjukan *stand-up comedy* Mens Rea pada Platform X didominasi oleh sentimen netral. Distribusi pada data hasil pelabelan akan ditampilkan pada Tabel 1.

Tabel 1. Sampel Pelabelan

No	Full Text	Label
1	@ainunnajib Mens rea nya jelas kok. Kuota haji dijual diam diam setelah ketahuan ada upaya pengembalian dana. Kenapa harus begitu? Kemana hasil penjualannya? Memangnya kuota haji itu milik pribadi atau negara?? Monggo @ainunnajib	*negative*
2	@meisjemetdparel Tidak ada mens rea hehe	neutral
3	@KangManto123 bener kata Pandji di mens rea knp suka bngt ya kita ini klo ada negosiasi psti rombongan sama kaya Elon Musk kmrn pas ke Indo	positive

4.2. *Preprocessing*

1. Cleaning

   Hasil proses *data cleaning* pada penelitian ini ditunjukkan pada Gambar 4.

	full_text	sentiment	tweet_clean
0	@ainunnajib Mens rea nya jelas kok. Kuota haji...	negative	Mens rea nya jelas kok Kuota haji dijual diam ...
1	@INAonTrending @KangManto123 Polisi usut ini k...	neutral	Polisi usut ini karena ada laporan resmi ke Po...
2	MENS REA stand-up comedy special Pandji Pragiw...	neutral	MENS REA stand up comedy special Pandji Pragiw...
3	@sudutkaca deym aman aja sih lebih wow mens re...	neutral	deym aman aja sih lebih wow mens rea soalnnya ...
4	@meisjemetdparel Tidak ada mens rea hehe	neutral	Tidak ada mens rea hehe
...	...	...	...
2845	@nuy_ladecima @pandji udah pernah nyoba dan se...	neutral	udah pernah nyoba dan semua pengalamannya udah...
2846	@grok @JukiHoki dalam kasus mens rea ini apa s...	neutral	dalam kasus mens rea ini apa sudah ada kerja s...
2847	Lagi rame dimana2 ngomongin Mens Rea tp sy leb...	positive	Lagi rame ngomongin Mens Rea tp sy lebih terta...
2848	@Fol1eadeux @ecommurz Kalo yg Mens Rea ttp gbs...	neutral	Kalo yg Mens Rea ttp gbs di download
2849	@muannas_alaidid udah mentok yaa cari topik yg...	negative	udah mentok yaa cari topik yg bisa masalahin d...

Gambar 4. Sampel Hasil Cleaning

Berdasarkan Gambar 4, proses *data cleaning* berhasil menghilangkan URL, *mention*, *hashtag*, emoji, tanda baca, angka, serta karakter yang tidak diperlukan sehingga data menjadi lebih bersih dan siap digunakan pada tahapan berikutnya.

2. Case Folding

   Hasil proses *case folding* ditunjukkan pada Gambar 5.

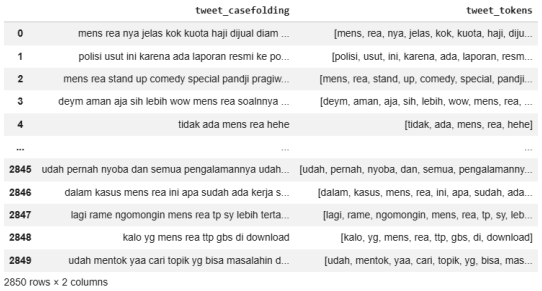
	full_text	tweet_casefolding
0	@ainunnajib Mens rea nya jelas kok. Kuota haji...	@ainunnajib mens rea nya jelas kok. kuota haji...
1	@INAonTrending @KangManto123 Polisi usut ini k...	@inaontrending @kangmanto123 polisi usut ini k...
2	MENS REA stand-up comedy special Pandji Pragiw...	mens rea stand-up comedy special pandji pragiw...
3	@sudutkaca deym aman aja sih lebih wow mens re...	@sudutkaca deym aman aja sih lebih wow mens re...
4	@meisjemetdparel Tidak ada mens rea hehe	@meisjemetdparel tidak ada mens rea hehe
5	@MisterBenComedy bukannya semenjak mens rea ri...	@misterbencomedy bukannya semenjak mens rea ri...
6	@kompascom Jadi ingat mens rea panji ada org y...	@kompascom jadi ingat mens rea panji ada org y...
7	@Fenrir3001 @RakaSwara25 @Otaku_Anime_Ind Di m...	@fenrir3001 @rakaswara25 @otaku_anime_ind di m...
8	masih sgt happy kalo inget kmrn nonton mens re...	masih sgt happy kalo inget kmrn nonton mens re...
9	@zainalamochtar @NataliusPigai2 produser tdk t...	@zainalamochtar @nataliuspigai2 produser tdk t...

Gambar 5. Sampel Hasil Case Folding

Berdasarkan Gambar 5, seluruh huruf pada tweet berhasil diubah menjadi huruf kecil (*lowercase*) sehingga penulisan kata menjadi lebih seragam dan mengurangi perbedaan bentuk penulisan.

3. Tokenizing

   Hasil proses *tokenizing* ditunjukkan pada Gambar 6.



	tweet_casefolding	tweet_tokens
0	mens rea nya jelas kok kuota haji dijual diam ...	[mens, rea, nya, jelas, kok, kuota, haji, diju...
1	polisi usut ini karena ada laporan resmi ke po...	[polisi, usut, ini, karena, ada, laporan, resm...
2	mens rea stand up comedy special pandji pragiw...	[mens, rea, stand, up, comedy, special, pandji...
3	deym aman aja sih lebih wow mens rea soalnnya ...	[deym, aman, aja, sih, lebih, wow, mens, rea, ...
4	tidak ada mens rea hehe	[tidak, ada, mens, rea, hehe]
...	...	...
2845	udah pernah nyoba dan semua pengalamannya udah...	[udah, pernah, nyoba, dan, semua, pengalamanny...
2846	dalam kasus mens rea ini apa sudah ada kerja s...	[dalam, kasus, mens, rea, ini, apa, sudah, ada...
2847	lagi rame ngomongin mens rea tp sy lebih terta...	[lagi, rame, ngomongin, mens, rea, tp, sy, leb...
2848	kalo yg mens rea ttp gbs di download	[kalo, yg, mens, rea, ttp, gbs, di, download]
2849	udah mentok yaa cari topik yg bisa masalahin d...	[udah, mentok, yaa, cari, topik, yg, bisa, mas...

2850 rows × 2 columns

Gambar 6. Sampel Hasil Tokenizing

Berdasarkan Gambar 6, setiap kalimat berhasil dipecah menjadi token berupa kata-kata tunggal sehingga memudahkan proses pengolahan data pada tahapan selanjutnya.

4. Normalization

Hasil proses *normalization* ditunjukkan pada Gambar 7.

```
Hasil spelling normalization:
                                        tweet_casefolding  \
0     mens rea nya jelas kok kuota haji dijual diam ...
1     polisi usut ini karena ada laporan resmi ke po...
2     mens rea stand up comedy special pandji pragiw...
3     deym aman aja sih lebih wow mens rea soalnnya ...
4                                tidak ada mens rea hehe
...                                                 ...
2845  udah pernah nyoba dan semua pengalamannya udah...
2846  dalam kasus mens rea ini apa sudah ada kerja s...
2847  lagi rame ngomongin mens rea tp sy lebih terta...
2848                kalo yg mens rea ttp gbs di download
2849  udah mentok yaa cari topik yg bisa masalahin d...
```

Gambar 7. Sampel Hasil Normalization

Berdasarkan Gambar 7, kata-kata tidak baku, singkatan, dan bahasa gaul berhasil diubah menjadi kata baku sehingga makna setiap tweet menjadi lebih konsisten.

5. Stopword Removal

Hasil proses *stopword removal* ditunjukkan pada Gambar 8.

	No	Sebelum Stopword	Sesudah Stopword
0	1	[polda, metro, jaya, menjadwalkan, pemeriksaan, klarifikasi, terhadap, pandji, pragiwaksono, pada, jumat, besok, terkait, materi, stand, up, mens, rea, polisi, akan, menggabungkan, lima, laporan, yang, masuk, menjadi, satu, sesi, pemeriksaan, agar, lebih, efisien, pandji, sendiri, memastikan, bakal, hadir, dan]	[polda, metro, menjadwalkan, pemeriksaan, klarifikasi, pandji, pragiwaksono, besok, terkait, stand, up, mens, rea, polisi, menggabungkan, laporan, masuk, sesi, pemeriksaan, efisien, pandji, hadir]
1	2	[kasus, dugaan, penistaan, agama, mens, rea, pandji, pragiwaksono, berlanjut, polisi, siapkan, ahli, kasus, yang, menyeret, nama, komika, pandji, pragiwaksono, berlanjut, pada, dugaan, penistaan, agama, dugaan, tersebut, muncul, dalam, laporan, dari, materi, stand, up, komedi, bertajuk, mens, rea, setelah]	[agama, mens, rea, pandji, pragiwaksono, berlanjut, polisi, siapkan, menyeret, komika, pandji, pragiwaksono, berlanjut, agama, muncul, laporan, stand, up, komedi, bertajuk, mens, rea]
2	3	[kamavg, berliut, rea, mens]	[kamavg, rea, mens]
3	4	[haha, ok, saja, kalau, begitu, share, dong, bit, favoritmu, yang, lain, dari, mens, rea, atau, stand, up, lain, atau, ada, komedian, indonesia, lain, yang, tajam, nyentil, politik]	[haha, favoritmu, mens, rea, stand, up, komedian, indonesia, tajam, nyentil, politik]
4	5	[mengakak, banget, hukum, di, negeri, ini, tidak, ada, mens, rea, katanya, terus, itu, duit, yang, digondol, buat, sedekah, apa, buat, ke, sokor, itu, hakim, nya, plenger, begitu]	[mengakak, hukum, negeri, mens, rea, duit, digondol, sedekah, sokor, hakim, plenger]
5	6	[mens, rea, panj, dliaporin, yang, begimian, saja, padahal, penghinaan, agama]	[mens, rea, panj, dliaporin, penghinaan, agama]

Gambar 8. Sampel Hasil stopword removal

Berdasarkan Gambar 8, kata-kata yang tidak memiliki kontribusi terhadap analisis sentimen berhasil dihilangkan sehingga hanya tersisa kata-kata yang memiliki informasi penting.

6. Stemming

Hasil proses *stemming* ditunjukkan pada Gambar 9.

	tweet_stopwords	tweet_stem
0	[mens, rea, kuota, haji, dijual, diam, diam, ketahuan, upaya, pengembalian, dana, kemana, hasil, penjualannya, memangnya, kuota, haji, pribadi, negara, monggo]	[mens, rea, kuota, haji, jual, diam, diam, tahu, upaya, kembali, dana, mana, hasil, jual, memang, kuota, haji, pribadi, negara, monggo]
1	[polisi, usut, laporan, resmi, polda, metro, stand, up, pandji, pragiwaksono, berjudul, mens, rea, tuduhannya, penodaan, agama, penghasutan, pasal, kuhp, dalami, penulis, skripnya, klarifikasi, unsur, pidana, panggungnya]	[polisi, usut, lapor, resmi, polda, metro, stand, up, pandji, pragiwaksono, judul, mens, rea, tuduh, noda, agama, hasut, pasal, kuhp, dalam, tulis, skrip, klarifikasi, unsur, pidana, panggung]
2	[mens, rea, stand, up, comedy, special, pandji, pragiwaksono, ditonton, keramaian, penasaran, komika, indonesia, berani, masuk]	[mens, rea, stand, up, comedy, special, pandji, pragiwaksono, tonton, ramai, penasaran, komika, indonesia, berani, masuk]
3	[deym, aman, wow, mens, rea, soalnnya, wkwkwk]	[deym, aman, wow, mens, rea, soalnnya, wkwkwk]
4	[mens, rea, hehe]	[mens, rea, hehe]
5	[semenjak, mens, rea, rilis, karir, told]	[semenjak, mens, rea, rilis, karir, told]
6	[mens, rea, panj, orang, stengahnya, stengah, jadikan, jabatan, jadikan, duit, sample]	[mens, rea, panj, orang, stengahnya, stengah, jadi, jabat, jadi, duit, sample]
7	[mens, rea, pandji, prawigaksono, bahas, prabowo, mengamankan, aktivis, komedi, serius]	[mens, rea, pandji, prawigaksono, bahas, prabowo, aman, aktivis, komedi, serius]

Gambar 9. Sampel Hasil stemming

Berdasarkan Gambar 9, kata-kata berimbuhan berhasil diubah menjadi bentuk kata dasar sehingga menghasilkan representasi teks yang lebih seragam sebelum dilakukan proses pembobotan TF-IDF.

4.3. Data Transformation

Sebelum dilakukan proses klasifikasi, dataset terlebih dahulu dibagi menjadi data latih (*training data*) dan data uji (*testing data*) menggunakan metode *train-test split*. Pada penelitian ini digunakan lima skenario pembagian data, yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50 untuk membandingkan kinerja model pada setiap proporsi data pelatihan dan pengujian. Setelah proses pembagian data selesai, dilakukan pembobotan kata menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) melalui TfidfVectorizer pada library Scikit-learn sehingga data teks dapat direpresentasikan ke dalam bentuk vektor numerik sebelum diproses menggunakan algoritma Naive Bayes. Pendekatan TF-IDF dipilih karena mampu memberikan bobot yang lebih tinggi pada kata-kata yang memiliki informasi penting sehingga dapat meningkatkan kualitas representasi fitur pada proses klasifikasi sentimen. Hal ini sejalan dengan penelitian yang dilakukan oleh Lestari dan Hutagalung [16] serta Karo et al. [17] yang menerapkan TF-IDF sebagai metode ekstraksi fitur sebelum proses klasifikasi menggunakan algoritma Naive Bayes. Berikut adalah contoh implementasi program Python Train dan Test

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer

# Membaca data
data = pd.read_csv(
    "hasil_preprocessing.csv",
    usecols=['sentiment', 'tweet_stem']
).dropna()

# Menampilkan data awal
print(data.head())

# Mengubah list menjadi kalimat
data['tweet_stem'] = data['tweet_stem'].str.replace(r"[\[\]']", "", regex=True)
data['tweet_stem'] = data['tweet_stem'].str.replace(",", "", regex=True)

# fitur dan label
X = data['tweet_stem']
y = data['sentiment'].astype(str)
# Fungsi untuk split dan TF-IDF
def prepare_data(X, y, test_size, random_state=42):

    # Split data
    X_train, X_test, y_train, y_test = train_test_split(
        X,
        y,
        test_size=test_size,
        stratify=y,
        random_state=random_state
    )

    # TF-IDF
    vectorizer = TfidfVectorizer(
        ngram_range=(1, 2),
        max_features=10000,
        min_df=2,
        sublinear_tf=True
    )

    X_train = vectorizer.fit_transform(X_train)
    X_test = vectorizer.transform(X_test)

    return X_train, X_test, y_train, y_test, vectorizer
```

Gambar 10.Metode TF-IDF

4.4. Data Mining

Tahap Data Mining dilakukan menggunakan algoritma Multinomial Naive Bayes untuk mengklasifikasikan data hasil pembobotan TF-IDF ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Model dibangun menggunakan data latih (*training data*) pada masing-masing skenario pembagian data, kemudian digunakan untuk memprediksi kelas sentimen pada data uji (*testing data*). Implementasi model dilakukan menggunakan library Scikit-learn dengan parameter alpha = 0,1. Pendekatan ini mengacu pada penelitian Hasanah dan Sari [15] yang menunjukkan bahwa algoritma Naive Bayes mampu memberikan performa yang baik pada proses klasifikasi sentimen berbasis teks.

```python
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, confusion_matrix

# Skenario 1 (90:10)
nb1 = MultinomialNB(alpha=0.1)
nb1.fit(X_train1, y_train1)
print('Accuracy Score Model (90:10):', round(accuracy_score(y_test1, nb1.predict(X_test1)), 4))

# Skenario 2 (80:20)
nb2 = MultinomialNB(alpha=0.1)
nb2.fit(X_train2, y_train2)
print('Accuracy Score Model (80:20):', round(accuracy_score(y_test2, nb2.predict(X_test2)), 4))

# Skenario 3 (70:30)
nb3 = MultinomialNB(alpha=0.1)
nb3.fit(X_train3, y_train3)
print('Accuracy Score Model (70:30):', round(accuracy_score(y_test3, nb3.predict(X_test3)), 4))

# Skenario 4 (60:40)
nb4 = MultinomialNB(alpha=0.1)
nb4.fit(X_train4, y_train4)
print('Accuracy Score Model (60:40):', round(accuracy_score(y_test4, nb4.predict(X_test4)), 4))

# Skenario 5 (50:50)
nb5 = MultinomialNB(alpha=0.1)
nb5.fit(X_train5, y_train5)
print('Accuracy Score Model (50:50):', round(accuracy_score(y_test5, nb5.predict(X_test5)), 4))
```

Gambar 11.Metode Naïve Bayes

Pada Tabel 2. Menampilkan Hasil Skenario yang dihasilkan pada penelitian.

Tabel 2. Hasil Skenario

Akurasi	Skenario
84%	Pertama
85%	Kedua
85%	Ketiga
84%	Keempat
85%	Kelima

Berdasarkan Tabel 2, skenario pertama dengan pembagian data 90:10 menghasilkan akurasi sebesar 84%. Pada skenario kedua dengan pembagian data 80:20, akurasi meningkat menjadi 85%. Skenario ketiga dengan pembagian data 70:30 juga memperoleh akurasi sebesar 85%. Selanjutnya, skenario keempat dengan pembagian data 60:40 menghasilkan akurasi sebesar 84%, sedangkan skenario kelima dengan pembagian data 50:50 kembali memperoleh akurasi sebesar 85%.

4.5. Evaluation

Hasil evaluasi model pada masing-masing skenario pembagian data disajikan pada subbab ini menggunakan *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

1 Skenario Pertama (90:10)

```
=================== SKENARIO 90:10 ====================
Model      : Multinomial Naive Bayes
Accuracy   : 0.8491

Confusion Matrix:
[[  1  18   0]
 [  2 241   2]
 [  0  21   0]]

Precision, Recall, F1-Score:
              precision    recall  f1-score   support

     Negatif     0.3333    0.0526    0.0909        19
      Netral     0.8607    0.9837    0.9181       245
     Positif     0.0000    0.0000    0.0000        21

    accuracy                         0.8491       285
   macro avg     0.3980    0.3454    0.3363       285
weighted avg     0.7621    0.8491    0.7953       285
```

Gambaar 12. Hasil confusion matrix scenario pertama

Berdasarkan hasil perhitungan metrik evaluasi yang diperoleh dari confusion matrix tersebut, nilai Precision pada kelas negatif adalah 0.3333, Recall sebesar 0.0526, dan F1-score sebesar 0.0909. Pada kelas netral, diperoleh nilai Precision sebesar 0.8607, Recall sebesar 0.9837, serta F1-score sebesar 0.9181. Sementara itu, pada kelas positif diperoleh nilai Precision,

Recall, dan F1-score masing-masing sebesar 0.0000. Hasil evaluasi tersebut mengindikasikan bahwa algoritma Multinomial Naïve Bayes mampu melakukan klasifikasi sentimen dengan cukup baik, khususnya pada kategori netral yang memperoleh nilai Precision, Recall, dan F1-score tertinggi dibandingkan kategori lainnya. Temuan ini menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengidentifikasi pola sentimen netral pada data tweet yang digunakan dalam penelitian.

2 Skenario Kedua (80:20)

```
=================== SKENARIO 80:20 ====================
Model      : Multinomial Naive Bayes
Accuracy   : 0.8579

Confusion Matrix:
[[  2  35   0]
 [  2 486   3]
 [  0  41   1]]

Precision, Recall, F1-Score:
              precision    recall  f1-score   support

     Negatif     0.5000    0.0541    0.0976        37
      Netral     0.8648    0.9898    0.9231       491
     Positif     0.2500    0.0238    0.0435        42

    accuracy                         0.8579       570
   macro avg     0.5383    0.3559    0.3547       570
weighted avg     0.7958    0.8579    0.8047       570

=======================================================
```

Gambaar 13. Hasil confusion matrix scenario kedua

Berdasarkan hasil perhitungan metrik evaluasi yang diperoleh dari confusion matrix tersebut, nilai Precision pada kelas negatif sebesar 0.5000, Recall sebesar 0.0541, dan F1-score sebesar 0.0976. Pada kelas netral diperoleh nilai Precision sebesar 0.8648, Recall sebesar 0.9898, serta F1-score sebesar 0.9231. Sementara itu, pada kelas positif diperoleh nilai Precision sebesar 0.2500, Recall sebesar 0.0238, dan F1-score sebesar 0.0435. Hasil evaluasi tersebut menunjukkan bahwa algoritma Multinomial Naïve Bayes mampu melakukan klasifikasi sentimen dengan cukup baik, khususnya pada kategori netral yang memperoleh nilai Precision, Recall, dan F1-score tertinggi dibandingkan kategori lainnya. Temuan tersebut mengindikasikan bahwa model memiliki kemampuan yang lebih baik dalam mengenali pola sentimen netral pada data tweet yang digunakan dalam penelitian ini.

3 Skenario Ketiga (70:30)

```
=================== SKENARIO 70:30 ====================
Model      : Multinomial Naive Bayes
Accuracy   : 0.8538

Confusion Matrix:
[[  1  55   0]
 [  1 728   7]
 [  0  62   1]]

Precision, Recall, F1-Score:
              precision    recall  f1-score   support

     Negatif     0.5000    0.0179    0.0345        56
      Netral     0.8615    0.9891    0.9209       736
     Positif     0.1250    0.0159    0.0282        63

    accuracy                         0.8538       855
   macro avg     0.4955    0.3410    0.3279       855
weighted avg     0.7836    0.8538    0.7971       855

=======================================================
```

Gambaar 14. Hasil confusion matrix scenario ketiga

Berdasarkan hasil perhitungan metrik evaluasi dari Confusion Matrix tersebut, diperoleh nilai Precision pada kelas negatif sebesar 0.5000, Recall sebesar 0.0179, dan F1-score sebesar 0.0345. Pada kelas netral diperoleh nilai Precision sebesar 0.8615, Recall sebesar 0.9891, serta F1-score sebesar 0.9209. Adapun pada kelas positif diperoleh nilai Precision sebesar 0.1250, Recall sebesar 0.0159, dan F1-score sebesar 0.0282. Nilai evaluasi yang diperoleh menunjukkan bahwa model Multinomial Naïve Bayes memberikan kinerja yang sangat baik pada klasifikasi sentimen netral, yang ditunjukkan oleh nilai Precision, Recall, dan F1-score yang lebih tinggi dibandingkan kategori sentimen lainnya. Namun demikian, model masih mengalami kendala dalam mengidentifikasi sentimen negatif dan positif karena sebagian besar data pada kedua kategori tersebut cenderung terprediksi sebagai sentimen netral.

4 Skenario Keempat (60:40)

```
=================== SKENARIO 60:40 ====================
Model      : Multinomial Naive Bayes
Accuracy   : 0.8509

Confusion Matrix:
[[  2  73   0]
 [  3 965  13]
 [  0  81   3]]

Precision, Recall, F1-Score:
              precision    recall  f1-score   support

     Negatif     0.4000    0.0267    0.0500        75
      Netral     0.8624    0.9837    0.9190       981
     Positif     0.1875    0.0357    0.0600        84

    accuracy                         0.8509      1140
   macro avg     0.4833    0.3487    0.3430      1140
weighted avg     0.7822    0.8509    0.7986      1140

=======================================================
```

Gambaar 15. Hasil confusion matrix scenario keempat

Berdasarkan hasil perhitungan metrik evaluasi yang diperoleh dari confusion matrix, nilai Precision pada kelas negatif sebesar 0.4000, Recall sebesar 0.0267, dan F1-score sebesar 0.0500. Pada kelas netral diperoleh nilai Precision sebesar 0.8624, Recall sebesar 0.9837, serta F1-score sebesar 0.9190. Sementara itu, pada kelas positif diperoleh nilai Precision sebesar 0.1875, Recall sebesar 0.0357, dan F1-score sebesar 0.0600. Berdasarkan nilai evaluasi tersebut, algoritma Multinomial Naïve Bayes menunjukkan kemampuan yang cukup baik dalam melakukan klasifikasi sentimen. Kinerja terbaik terlihat pada kelas netral yang memperoleh nilai Precision, Recall, dan F1-score lebih tinggi dibandingkan kelas lainnya. Kondisi tersebut menunjukkan bahwa model lebih efektif dalam mengidentifikasi pola sentimen netral pada data tweet yang digunakan dalam penelitian ini.

5 Skenario kelima (50:50)

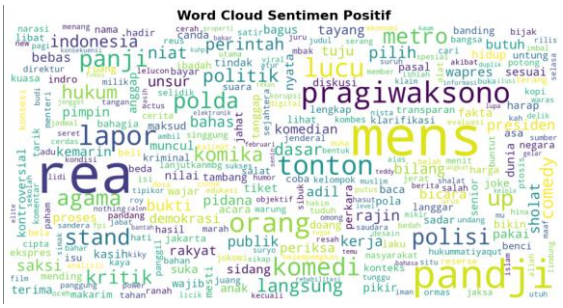
```
==================== SKENARIO 50:50 ====================
Model      : Multinomial Naive Bayes
Accuracy   : 0.8547

Confusion Matrix:
[[   1   92    0]
 [   4 1214    9]
 [   0  102    3]]

Precision, Recall, F1-Score:
              precision    recall  f1-score   support

     Negatif     0.2000    0.0108    0.0204        93
      Netral     0.8622    0.9894    0.9214      1227
     Positif     0.2500    0.0286    0.0513       105

    accuracy                         0.8547      1425
   macro avg     0.4374    0.3429    0.3310      1425
weighted avg     0.7739    0.8547    0.7985      1425

============================================================
```

Gambaar 16. Hasil confusion matrix scenario kelima

Berdasarkan hasil perhitungan metrik evaluasi yang diperoleh dari confusion matrix tersebut, nilai Precision pada kelas negatif sebesar 0.2000, Recall sebesar 0.0108, dan F1-score sebesar 0.0204. Pada kelas netral diperoleh nilai Precision sebesar 0.8622, Recall sebesar 0.9894, serta F1-score sebesar 0.9214. Sementara itu, pada kelas positif diperoleh nilai Precision sebesar 0.2500, Recall sebesar 0.0286, dan F1-score sebesar 0.0513. Hasil pengujian menunjukkan bahwa algoritma Multinomial Naïve Bayes masih mampu memberikan kinerja yang baik dalam proses klasifikasi sentimen. Performa terbaik ditunjukkan oleh kelas netral yang memperoleh nilai Precision, Recall, dan F1-score paling tinggi dibandingkan kelas lainnya. Kondisi tersebut mengindikasikan bahwa model lebih efektif dalam mengenali karakteristik sentimen netral pada data tweet yang digunakan dalam penelitian ini.

4.6. Visualisai Wordcloud

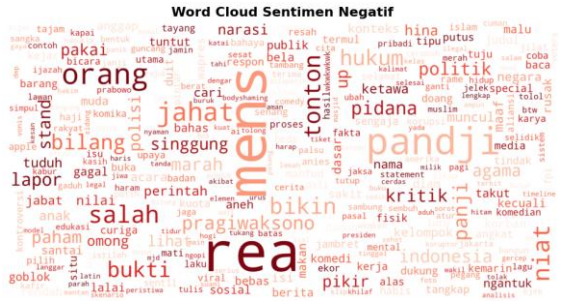
Visualisasi *WordCloud* digunakan untuk menampilkan kata-kata yang paling dominan berdasarkan frekuensi kemunculannya pada setiap kategori sentimen. Pada penelitian ini, visualisasi *WordCloud* dibuat untuk masing-masing kelas sentimen, yaitu positif, negatif, dan netral, sehingga dapat memberikan gambaran mengenai kata-kata yang paling sering muncul pada setiap kelompok sentiment.



Gambar 17. Wordcloud sentimen positif

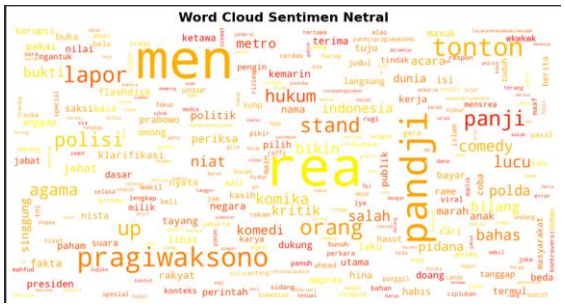
Visualisasi WordCloud sentimen positif ditunjukkan pada Gambar 17. Pada visualisasi tersebut, beberapa kata seperti “pandji”, “mens”, “komedi”, “lucu”, “kreatif”, dan “agama” muncul dengan ukuran yang relatif lebih besar dibandingkan kata lainnya. Dominasi kata-kata tersebut menunjukkan bahwa opini positif pengguna media sosial cenderung berkaitan dengan apresiasi terhadap materi komedi, gaya penyampaian, serta pembahasan isu sosial dalam pertunjukan stand-up comedy Mens Rea. Selain itu, kemunculan kata seperti “demokrasi”, “komunikasi”, dan “berani” menggambarkan adanya tanggapan positif terhadap penyampaian kritik sosial dan opini publik yang dibawakan dalam pertunjukan tersebut. Pola kemunculan kata-kata tersebut menunjukkan bahwa sentimen positif pada dataset penelitian dipengaruhi oleh pandangan masyarakat yang menilai pertunjukan stand-up comedy Mens Rea tidak hanya sebagai hiburan, tetapi juga

sebagai media penyampaian kritik dan diskusi sosial.



Gambar 18. Wordcloud sentimen Negatif

Visualisasi WordCloud sentimen negatif ditunjukkan pada Gambar 18. Pada visualisasi tersebut, beberapa kata seperti “pandji”, “hina”, “salah”, “agama”, “kritik”, dan “pidana” muncul dengan ukuran yang relatif lebih besar dibandingkan kata lainnya. Dominasi kata-kata tersebut menunjukkan bahwa opini negatif pengguna media sosial cenderung berkaitan dengan kritik dan ketidaksetujuan terhadap materi yang disampaikan dalam pertunjukan stand-up comedy Mens Rea. Selain itu, kemunculan kata seperti “marah”, “lapor”, “singgung”, dan “rusak” menggambarkan adanya tanggapan negatif terhadap isu yang dianggap sensitif oleh sebagian pengguna media sosial. Pola kemunculan kata-kata tersebut menunjukkan bahwa sentimen negatif pada dataset penelitian dipengaruhi oleh respons masyarakat terhadap materi pertunjukan yang menimbulkan kontroversi dan perdebatan di media sosial.



Gambar 19. Wordcloud sentimen Netral

Visualisasi WordCloud sentimen netral ditunjukkan pada Gambar 19. Pada visualisasi tersebut, beberapa kata seperti “pandji”, “mens”, “komedi”, “agama”, “politik”, dan “acara” muncul dengan ukuran yang relatif lebih besar dibandingkan kata lainnya. Dominasi kata-kata tersebut menunjukkan bahwa sebagian besar percakapan pengguna media sosial pada kategori netral lebih banyak berkaitan dengan penyampaian informasi dan pembahasan umum terkait pertunjukan stand-up comedy Mens Rea. Selain itu, kemunculan kata seperti “opini”, “materi”, dan “publik” menggambarkan bahwa percakapan pada kategori sentimen netral cenderung berfokus pada isi pertunjukan serta isu yang dibahas tanpa menunjukkan kecenderungan sentimen positif maupun negatif secara dominan. Pola kemunculan kata-kata tersebut menunjukkan bahwa sentimen netral pada dataset penelitian dipengaruhi oleh banyaknya pengguna media sosial yang memberikan tanggapan informatif dan diskusi umum terkait pertunjukan stand-up comedy Mens Rea di media sosial X.

5. KESIMPULAN

Pada penelitian ini telah berhasil dalam mengimplementasikan metode TF-IDF dan algoritma Multinomial Naïve Bayes untuk menganalisis sentimen opini publik terhadap pertunjukan stand-up comedy Mens Rea pada Platform X. Berdasarkan pengujian pada lima skenario pembagian data (90:10, 80:20, 70:30, 60:40, dan 50:50), model menghasilkan performa klasifikasi yang relatif stabil dengan akurasi berkisar antara 84,91%–85,79%. Skenario pembagian data 80:20 memberikan performa terbaik dengan akurasi sebesar 85,79%. Selain itu, hasil evaluasi menunjukkan bahwa model lebih optimal dalam mengklasifikasikan sentimen netral, sedangkan performa pada kelas positif dan negatif masih dipengaruhi oleh ketidakseimbangan distribusi data. Temuan ini menunjukkan bahwa kombinasi TF-IDF dan Multinomial Naïve Bayes efektif digunakan untuk analisis sentimen teks berbahasa Indonesia serta dapat menjadi dasar pengembangan penelitian selanjutnya dengan menerapkan teknik penyeimbangan data atau optimasi model guna meningkatkan performa klasifikasi pada seluruh kelas sentimen.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan apresiasi kepada seluruh pihak yang telah membantu dan memberikan dukungan, arahan, serta kontribusi selama dalam proses penelitian dan penyusunan

pada artikel ini sehingga penelitian dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

[1] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, 'The echo chamber effect on social media', *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, Mar. 2021, doi: 10.1073/pnas.2023301118.

[2] Y. K. Dwivedi, E. Ismagilova, N. P. Rana, and R. Raman, 'Social Media Adoption, Usage And Impact In Business-To-Business (B2B) Context: A State-Of-The-Art Literature Review', *Information Systems Frontiers*, vol. 25, no. 3, pp. 971–993, Jun. 2023, doi: 10.1007/s10796-021-10106-y.

[3] M. Wankhade, A. C. S. Rao, and C. Kulkarni, 'A survey on sentiment analysis methods, applications, and challenges', *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[4] W. Ahmed, A. Fenton, M. Hardey, and R. Das, 'Binge Watching and the Role of Social Media Virality towards promoting Netflix's Squid Game', *IIM Kozhikode Society & Management Review*, vol. 11, no. 2, pp. 222–234, Jul. 2022, doi: 10.1177/22779752221083351.

[5] R. Wati, S. Ernawati, and H. Rachmi, 'Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH', *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, Apr. 2023, doi: 10.34010/jamika.v13i1.9424.

[6] S. Zervoudakis, E. Marakakis, H. Kondylakis, and S. Goumas, 'OpinionMine: A Bayesian-based framework for opinion mining using Twitter Data', *Machine Learning with Applications*, vol. 3, p. 100018, Mar. 2021, doi: 10.1016/j.mlwa.2020.100018.

[7] D. Kusnanda and A. Permana, 'Implementation of Naive Bayes Classifier (NBC) for Sentiment Analysis on Twitter in Mobile Legends', *International Journal of Science, Technology & Management*, vol. 4, no. 5, pp. 1132–1138, Sep. 2023, doi: 10.46729/ijstm.v4i5.935.

[8] M. Wankhade, A. C. S. Rao, and C. Kulkarni, 'A survey on sentiment analysis methods, applications, and challenges', *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[9] A. Muzaki and A. Witanti, 'Sentiment Analysis Of The Community In The Twitter To The 2020 Election In Pandemic Covid-19 By Method Naive Bayes Classifier', *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.

[10] B. Liu, 'Sentiment Analysis: A Fascinating Problem', in *Sentiment Analysis and Opinion Mining*, 2012, pp. 1–8. doi: 10.1007/978-3-031-02145-9_1.

[11] V. B. Lestari and C. A. Hutagalung, 'Data Mining Approach: K-Means Clustering and Naïve Bayes Classifier for Graduate Quality Analysis', *J-KOMA : Jurnal Ilmu Komputer dan Aplikasi*, vol. 8, no. 1, pp. 33–42, Jun. 2025, doi: 10.21009/j-koma.v8i1.05.

[12] A. N. Hasanah and B. N. Sari, 'Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naïve Bayes Classifier', *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, pp. 90–96, Jan. 2024, doi: 10.23960/jitet.v12i1.3628.

[13] M. Jemboin, D. M. D. Utami Putra, I. G. M. Yudi Antara, I. G. A. Indrawan, and I. G. I. Sudipa, 'Public Opinion Sentiment Analysis of the #BoikotTrans7 Hashtag Case on YouTube Using the Naive Bayes Classifier Method', *Journal of Business, Social and Technology*, vol. 7, no. 3, pp. 1–11, Jun. 2026, doi: 10.59261/jbt.v7i3.673.

[14] O. D. Palupi, Y. K. Sari, and J. Iskandar, 'Comparison of CNN and SVM Algorithms in Sentiment Analysis of Roblox Game Based on Bug, Loading Asset, and Connection Stability Aspects', *G-Tech: Jurnal Teknologi Terapan*, vol. 10, no. 2, pp. 710–719, Apr. 2026, doi: 10.70609/g-tech.v10i2.9156.

[15] M. Jemboin, D. M. D. Utami Putra, I. G. M. Yudi Antara, I. G. A. Indrawan, and I. G. I. Sudipa, 'Public Opinion Sentiment Analysis of the #BoikotTrans7 Hashtag Case on YouTube Using the Naive Bayes Classifier Method', *Journal of Business, Social and Technology*, vol. 7, no. 3, pp. 1–11, Jun. 2026, doi: 10.59261/jbt.v7i3.673.