

ANALISIS SENTIMEN TERHADAP PEMBERITAAN HASIL REKAPITULASI PEMILU PRESIDEN 2024 PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN NAIVE BAYES CLASSIFIER

M.Riswan^{1*}, Aji Primajaya², Agung Susilo Yuda Irawan³

^{1,2,3}I Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; telp (0267)641177

Received: 19 November 2024

Accepted: 14 Januari 2025

Published: 20 Januari 2025

Keywords:

Sentiment Analysis;

Naïve Bayes;

Instagram;

Knowledge Discovery in Database (KDD).

Correspondent Email:

2010631170017@student.u

nsika.ac.id

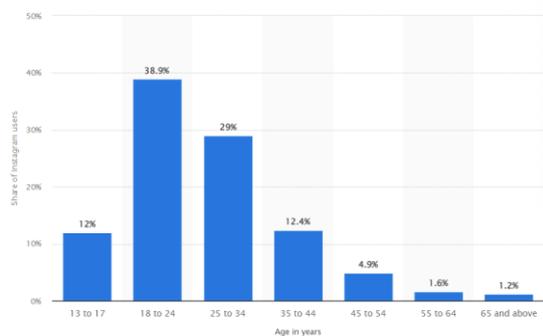
Abstrak. Media sosial Instagram, dengan lebih dari 100 juta pengguna di Indonesia yang didominasi kelompok usia produktif, telah menjadi platform utama dalam penyebaran informasi dan pembentukan opini publik terkait hasil rekapitulasi pemilu presiden 2024. Tantangan muncul terkait penyebaran informasi yang belum terverifikasi dan polarisasi pendapat yang dapat mempengaruhi stabilitas politik pasca pemilu. Penelitian ini bertujuan untuk mengidentifikasi dan mengkategorikan sentimen masyarakat terhadap hasil rekapitulasi pemilu presiden 2024 di Instagram serta menerapkan metode naive bayes untuk klasifikasi sentimen. Metodologi Knowledge Discovery in Database (KDD) diterapkan melalui tahapan data Selection, preprocessing, transformation, data mining dengan naive bayes classifier, dan evaluation. Data diambil sekitar 2000 komentar pada unggahan berita terkait hasil rekapitulasi pemilu di akun media massa terpercaya seperti kompastv, metrotv, dan tvonenews menggunakan tools scraping phantombuster. Hasil evaluasi menunjukkan algoritma naive bayes classifier memiliki rata-rata accuracy 75%, precision 75%, recall 74%, dan F1-score 74%, dengan skenario pembagian data 90:10 memberikan performa terbaik (accuracy 76%, precision 77%, recall 75%, F1-score 75%). Analisis sentimen mengungkapkan dominasi sentimen negatif sebesar 54,4% dalam komentar terkait hasil Pemilu 2024. Temuan ini bermanfaat untuk evaluasi respons masyarakat dan mendukung pengambilan keputusan yang lebih tepat oleh pemerintah dan lembaga terkait.

Abstract. Instagram, with over 100 million users in Indonesia, predominantly from the productive age group, has become a key platform for the dissemination of information and shaping public opinion regarding the 2024 presidential election results. Challenges arise from the spread of unverified information and polarized opinions that may affect post-election political stability. This study aims to identify and categorize public sentiment towards the 2024 presidential election result recap on Instagram, applying the Naive Bayes method for sentiment classification. The Knowledge Discovery in Database (KDD) methodology was applied, including data selection, preprocessing, transformation, data mining with Naive Bayes classifier, and evaluation. Data from around 2,000 comments on posts from trusted media accounts like KompasTV, MetroTV, and TVOneNews were collected using PhantomBuster scraping tools. Evaluation results show an average accuracy

of 75%, precision of 75%, recall of 74%, and F1-score of 74%, with the 90:10 data split scenario performing the best (accuracy 76%, precision 77%, recall 75%, F1-score 75%). Sentiment analysis revealed a dominance of negative sentiment at 54.4% in comments about the 2024 election results. These findings are valuable for evaluating public response and supporting better decision-making by the government and related institutions.

1. PENDAHULUAN

Pemilihan umum (pemilu) merupakan momentum krusial dalam perjalanan demokrasi suatu negara, khususnya dalam menentukan kepemimpinan bangsa ke depan. Hasil rekapitulasi Pemilu Presiden 2024 telah menjadi fokus perhatian publik, terutama di ranah digital yang kini memiliki pengaruh signifikan terhadap pembentukan opini publik. Platform media sosial Instagram, dengan penetrasi lebih dari 100 juta pengguna di Indonesia, telah menjadi saluran utama diskusi dan penyebaran informasi. Dominasi pengguna Instagram berada pada kelompok usia 18-24 tahun (38,9%) dan 25-34 tahun (29%), menegaskan perannya sebagai medium penting untuk menjangkau mayoritas penduduk usia produktif di Indonesia [1], seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Pengguna Instagram Rentang Usia Tahun 2023

Namun, sejumlah tantangan muncul terkait pemberitaan hasil rekapitulasi Pemilu Presiden 2024 di media sosial. Penyebaran informasi yang belum terverifikasi dan polarisasi pendapat yang tajam berpotensi mempengaruhi pandangan masyarakat serta memicu ketegangan sosial. Situasi ini dipersulit dengan adanya kesulitan dalam mengidentifikasi dan mengukur sentimen publik secara akurat dan cepat.

Penelitian terdahulu oleh Fitriyyah telah menunjukkan efektivitas metode Naive Bayes dalam analisis sentimen politik di Indonesia,

dengan tingkat akurasi mencapai 88% untuk klasifikasi dua kelas [2]. Namun, masih terdapat celah dalam literatur terkait analisis sentimen terhadap hasil pemilu presiden terbaru di Indonesia, khususnya di platform Instagram.

Penelitian ini bertujuan untuk mengidentifikasi dan mengkategorikan sentimen masyarakat terhadap hasil rekapitulasi Pemilu Presiden 2024 di Instagram menggunakan metode Naive Bayes Classifier. Data penelitian diperoleh dari sekitar 2000 komentar pada unggahan berita terkait hasil rekapitulasi Pemilu di akun-akun media massa terpercaya. Pengumpulan data dilakukan menggunakan tools scraping seperti Phantombuster.

Melalui penelitian ini, diharapkan dapat diperoleh wawasan komprehensif mengenai sentimen masyarakat terhadap hasil Pemilu 2024, yang dapat mendukung pengambilan keputusan yang lebih tepat oleh pemerintah, penyelenggara pemilu, dan pihak-pihak terkait lainnya. Hasil penelitian juga berkontribusi dalam pengembangan metode analisis sentimen yang lebih akurat untuk konteks politik Indonesia.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen atau opinion mining adalah cabang ilmu yang mempelajari dan menganalisis opini, perasaan, evaluasi, sikap, emosi dan pandangan masyarakat terhadap suatu entitas seperti produk, jasa, organisasi, individu, isu, peristiwa, topik dan atributnya yang diekspresikan dalam bentuk teks [3].

2.2 Pemilu

Pemilihan Umum (Pemilu) adalah instrumen penting dalam demokrasi yang mewadahi kedaulatan rakyat untuk memilih wakil-wakil mereka di berbagai tingkat pemerintahan, termasuk anggota DPR, DPD, Presiden dan Wakil Presiden, serta DPRD [4].

Dalam pelaksanaannya, pemilu harus memenuhi asas langsung, umum, bebas, rahasia, jujur, dan adil untuk menjamin legitimasi hasil pemilihan dan kepercayaan publik terhadap proses demokratis.

2.3 *Crawling Data*

Crawling data merupakan teknik otomatisasi untuk mengumpulkan informasi dari berbagai sumber digital, terutama dari internet dan media sosial. Dalam era digital saat ini, proses *crawling* telah berkembang dari pendekatan berbasis kode menjadi solusi berbasis *tools* yang lebih mudah digunakan dan efisien. *Tools crawling* modern menyediakan antarmuka yang lebih *user-friendly* dan fitur otomatisasi yang memungkinkan pengguna untuk mengekstrak data tanpa perlu memahami bahasa pemrograman secara mendalam. Proses ini tetap harus memperhatikan batasan *API (Application Programming Interface)* dan kebijakan *privasi platform* terkait, serta aspek etika pengumpulan data untuk memastikan kepatuhan terhadap standar privasi dan keamanan data [5].

2.4 *Data Mining*

Data mining merupakan suatu proses sistematis untuk pengumpulan, pengolahan, dan analisis data yang bertujuan mengekstrak informasi penting dan berharga dari kumpulan data yang besar. Proses tersebut dapat dicapai melalui implementasi perangkat lunak yang mengintegrasikan penghitungan matematika atau teknik kecerdasan buatan untuk menemukan pola dan hubungan yang bermakna dalam data. Dalam penerapannya, data mining memiliki beberapa metodologi standar, di antaranya yang paling umum digunakan adalah SEMMA (*Sample* untuk pengambilan sampel data, *explore* untuk eksplorasi data, *modify* untuk modifikasi dan transformasi data, model untuk pembuatan model analisis, dan *assess* untuk evaluasi hasil) [6]. Hal tersebut memungkinkan analisis data secara terstruktur dan komprehensif untuk menghasilkan wawasan yang dapat digunakan dalam pengambilan keputusan.

2.5 *Knowledge Discovery in Databases (KDD)*

Knowledge Discovery in Databases (KDD) merupakan proses sistematis untuk

mengidentifikasi pola yang valid, baru, berpotensi bermanfaat, dan dapat dipahami dari kumpulan data yang kompleks [7]. Proses KDD terdiri dari beberapa tahapan krusial, meliputi *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*, yang secara keseluruhan bertujuan menghasilkan pengetahuan yang bernilai dari data mentah, berikut tahapan – tahapannya.

1. *Data Selection*, mengumpulkan data yang relevan dari berbagai sumber, seperti media sosial, basis data internal, atau berita.
2. *Preprocessing*, membersihkan dan menyiapkan data dengan menghapus noise serta elemen yang tidak relevan.
3. *Transformation*, mengubah data ke dalam bentuk yang sesuai untuk proses data mining, misalnya melalui ekstraksi fitur atau normalisasi.
4. *Data Mining*, mengidentifikasi pola dalam data menggunakan algoritma seperti klasifikasi dan klustering.
5. *Evaluation*, mengevaluasi hasil dari data mining untuk memastikan bahwa pola atau pengetahuan yang ditemukan memiliki nilai dan relevansi yang signifikan.

2.6 *Naïve Bayes*

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang sederhana namun efektif untuk klasifikasi teks, termasuk analisis sentimen. Algoritma ini didasarkan pada *Teorema Bayes*, yang menghitung probabilitas suatu kelas berdasarkan fitur yang muncul dalam data [8]. Formula dari *Teorema Bayes* untuk memprediksi kelas H dengan data X diberikan sebagai berikut.

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

Dimana:

- $P(H|X)$: Probabilitas kelas H diberikan data X (probabilitas posterior).
- $P(X|H)$: Probabilitas kemunculan data X dalam kelas H (*likelihood*).
- $P(H)$: Probabilitas awal dari kelas H (prior).
- $P(X)$: Probabilitas awal dari data X .

2.7 Preprocessing

Preprocessing adalah tahapan penting dalam *text mining* untuk mengubah data tidak terstruktur menjadi data terstruktur yang siap untuk dianalisis. Tahapan *preprocessing* meliputi beberapa langkah seperti *cleaning*, *case folding*, *stopword removal*, *tokenizing*, dan *stemming*. Tahapan-tahapan ini membantu dalam menyederhanakan data dan meningkatkan kualitas hasil analisis sentimen [9].

1. *Cleaning*, menghilangkan noise atau data yang tidak relevan, seperti URL, simbol, dan karakter khusus.
2. *Case Folding*, mengonversi semua teks menjadi huruf kecil untuk mengurangi variasi kata yang tidak diperlukan.
3. *Stopword Removal*, menghapus kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis, seperti "dan", "yang", atau "dari".
4. *Tokenizing*, memecah teks menjadi bagian kecil yang disebut token, seperti kata atau frase, untuk mempermudah analisis lebih lanjut.
5. *Stemming*, mengubah kata-kata ke bentuk dasarnya, sehingga kata dengan makna serupa namun bentuk berbeda dapat disatukan, misalnya, "berlari", "berlarian" menjadi "lari".

2.8 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (*TF-IDF*) adalah metode pembobotan kata yang mengkombinasikan frekuensi kemunculan kata dalam dokumen (*TF*) dengan *inverse* dari frekuensi dokumen yang mengandung kata tersebut (*IDF*) [10]. Metode ini efektif dalam mengidentifikasi kata-kata yang paling relevan dalam suatu dokumen dengan memberikan bobot lebih tinggi pada kata yang sering muncul dalam dokumen tertentu namun jarang ditemukan di dokumen lain.

Dalam analisis sentimen media sosial, *TF-IDF* berperan penting dalam mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Metode ini terbukti efektif dalam meningkatkan akurasi klasifikasi sentimen dengan mempertimbangkan konteks dan kepentingan relatif setiap kata dalam dataset [11].

2.9 Confusion Matrix

Confusion Matrix adalah metode evaluasi yang digunakan untuk mengukur performa model klasifikasi dalam machine learning dengan menyajikan hasil prediksi dalam bentuk matriks [12]. Matriks ini membandingkan hasil prediksi model dengan nilai aktual (*ground truth*) melalui empat komponen utama: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*.

Dalam analisis sentimen yang membagi sentimen menjadi dua kategori (positif dan negatif), *confusion matrix* berfungsi untuk menunjukkan jumlah data yang diprediksi dengan benar serta yang diprediksi keliru pada setiap kategori [13]. Ada tiga elemen utama dalam confusion matrix.

1. Akurasi (*Accuracy*) mengukur persentase prediksi yang benar, baik untuk kelas positif maupun negatif, terhadap total keseluruhan data. Akurasi dapat dihitung dengan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision* menghitung proporsi data yang diprediksi positif dengan benar dibandingkan dengan semua data yang diprediksi sebagai positif. *Precision* dihitung dengan formula:

$$Precision = \frac{TP}{TP + FP}$$

3. *Recall* mengukur persentase data positif yang diprediksi dengan benar dibandingkan dengan seluruh data positif yang ada. *Recall* dirumuskan sebagai:

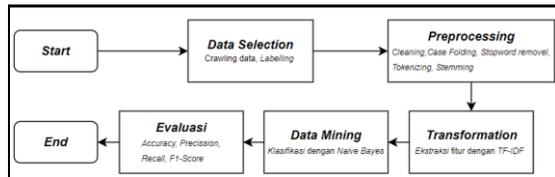
$$Recall = \frac{TP}{TP + FN}$$

Keterangan:

- *TP* = Jumlah data yang diprediksi positif dan memang benar positif.
- *TN* = Jumlah data yang diprediksi negatif dan memang benar negatif.
- *FP* = Jumlah data yang diprediksi positif tetapi sebenarnya negatif.
- *FN* = Jumlah data yang diprediksi negatif tetapi sebenarnya positif.

3. METODE PENELITIAN

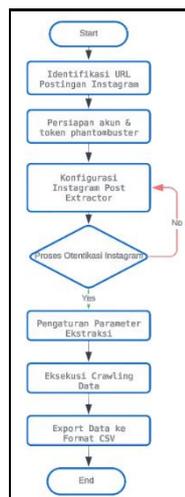
Penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases (KDD)* untuk menganalisis sentimen terkait pemberitaan hasil rekapitulasi pemilu Presiden 2024 di Instagram. Alur penelitian ini mencakup beberapa tahap, yaitu data selection, preprocessing, transformation, data mining, dan evaluation yang dapat dilihat pada Gambar 2.



Gambar 2. Alur Penelitian

3.1. Data Selection

Data yang digunakan dalam penelitian ini berupa komentar teks pada unggahan terkait hasil rekapitulasi Pemilu Presiden 2024 dari akun media massa seperti @kompastv, @metrotv, @tvonenews dan beberapa media massa lainnya. Data dikumpulkan menggunakan teknik web scraping dengan bantuan alat otomatisasi Phantombuster yang disimpan dalam format CSV berisi informasi seperti nama pengguna, teks komentar, dan waktu komentar. Pengumpulan data dilakukan pada 21 Februari 2024, dan hanya mencakup akun terverifikasi yang memenuhi kriteria tertentu, berikut ini tahapan – tahapannya pada Gambar 3.



Gambar 3. Flowchart Crawling Data

Setelah data terkumpul, proses pelabelan dilakukan dengan memberikan label "Positif" atau "Negatif" pada setiap komentar. Untuk efisiensi, data teks diterjemahkan ke bahasa

Inggris dan diberi label secara otomatis dengan bantuan alat *machine learning*.

3.2. Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan dan menyusun data sehingga siap untuk analisis lebih lanjut. Langkah-langkah *preprocessing* mencakup.

1. *Cleaning*, menghapus elemen yang tidak relevan seperti emotikon, hashtag, URL, dan nama pengguna.
2. *Case Folding*, mengubah semua teks menjadi huruf kecil untuk konsistensi.
3. *Tokenizing*, memecah teks menjadi kata-kata atau token.
4. *Stopword Removal*, menghapus kata-kata umum yang tidak berkontribusi signifikan.
5. *Stemming*, mengubah kata ke bentuk dasarnya untuk menyatukan variasi kata.

3.3. Transformation

Data komentar yang telah melalui tahap preprocessing kemudian diubah ke dalam bentuk numerik menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)*. *TF-IDF* memberikan bobot pada setiap kata berdasarkan frekuensi dan keunikannya di seluruh kumpulan data. Bobot ini menjadi representasi numerik yang siap digunakan sebagai input untuk model klasifikasi.

3.4. Data Mining

Tahap *data mining* bertujuan untuk melakukan klasifikasi sentimen pada komentar Instagram menggunakan algoritma *naive bayes classifier*. Setelah data disiapkan melalui *preprocessing* dan pembobotan *TF-IDF*, model *naive bayes* digunakan untuk membedakan antara sentimen positif dan negatif. Model dilatih menggunakan data pelatihan, kemudian diuji dengan data pengujian untuk menilai kemampuannya dalam mengklasifikasikan sentimen secara akurat.

3.5. Evaluation

Evaluasi model dilakukan untuk mengukur keakuratan klasifikasi. Matriks konfusi (*confusion matrix*) digunakan untuk menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, yang memberikan gambaran tentang performa model. Metrik ini

dibandingkan antara hasil prediksi dan data aktual untuk menilai kinerja model *naive bayes* dalam klasifikasi sentimen terkait hasil rekapitulasi pemilu 2024 di Instagram.

4. HASIL DAN PEMBAHASAN

Penelitian ini menganalisis sentimen masyarakat terhadap pemberitaan hasil rekapitulasi Pemilu Presiden 2024 di media sosial Instagram menggunakan algoritma naive bayes classifier. Proses penelitian mencakup tahapan pengumpulan data komentar, preprocessing, transformasi data, klasifikasi sentimen, dan evaluasi.

4.1 Data Selection

Data dikumpulkan dari komentar pada unggahan beberapa akun media massa pada tanggal 21 Februari 2024. Pengumpulan dilakukan dengan bantuan tool phantomBuster, melalui tahapan berikut:

1. Identifikasi UR, menelusuri URL unggahan yang relevan.
2. Persiapan Akun & Token, verifikasi akun Phantombuster untuk mendapatkan API token.
3. Konfigurasi Ekstraksi, mengatur parameter seperti jumlah komentar per unggahan dan jenis data yang diambil (*username*, teks, *timestamp*).
4. Proses Otentikasi, *login* Instagram untuk otorisasi penuh.
5. Eksekusi *Crawling*, pengambilan data komentar secara otomatis.
6. Ekspor Data, data disimpan dalam format CSV untuk proses selanjutnya.

Setelah pengumpulan, data dievaluasi dan dilakukan normalisasi ejaan agar sesuai dengan standar KBBI. Data akhir yang digunakan berjumlah 2.194 komentar Gambar 4.

username	comment	commentDate	comments	label
m_kaculbnyah	01 & 03 Ilmu token listrik, semesta sedikit s...	2024-02-20T16:03:51.000Z	0/1 & 0/2 electric token science, the less the...	NEGATIVE
cahya_kristiana	Alhamdulillah abah selamat amanah berat	2024-02-20T15:38:08.000Z	hAlhamdulillah i was scared from a heavy load	NEGATIVE
rela_rafiani	Kita tinggal hai saja dengan janji2nya aman...	2024-02-20T15:41:42.000Z	hWe've just have to see whether the promise...	NEGATIVE
deni_hermani_sakaria	Abah tidak tidur, kebesaran akan segera lek...	2024-02-20T15:08:12.000Z	hAbah does not sleep, the truth will soon b...	NEGATIVE
alfans	Angka itu sudah tidak penting bagi saya prib...	2024-02-20T15:27:52.000Z	hThat figure is no longer important to me per...	NEGATIVE

(Gambar 4. Data Mentah)

4.2 Preprocessing

Tahap *preprocessing* dilakukan untuk menghilangkan elemen yang tidak relevan dan

menyiapkan data untuk analisis. Tahapan ini meliputi:

1. *Cleaning*, menghapus emotikon, URL, mention, dan simbol.
2. *Case Folding*, mengonversi semua huruf menjadi huruf kecil.
3. *Tokenizing*, memecah kalimat menjadi kata-kata individual.
4. *Stopword Removal*, menghilangkan kata-kata umum yang tidak signifikan.
5. *Stemming*, mengubah kata-kata ke bentuk dasar.

Hasil dari preprocessing dapat dilihat pada gambar dibawah ini (Gambar 5).

	stemming
0	ilmu token listrik makin sedikit makin berisik
1	alhamdulillah abah selamat amanah berat
2	tinggal liat dengan janji amanah dan siap paja...
3	allah tidur benar segera kuat siap lah hadap a...
4	angka penting saya pribadi awal sudah tidak be...
...	...
1952	alhamdulillah sesuai prediksi moga konstitusi ...
1953	komentar hancur komentar ayo sama bangun senan...
1954	jokowi periode second account
1955	sabar liat gibran ngomong d sidang pbb terus c...
1956	capres nya langgar ham wakil presiden nya lang...

(Gambar 5. Data Hasil Preprocessing)

4.3 Transformation

Proses transformasi menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) untuk memberikan bobot pada kata berdasarkan frekuensi dan keunikannya. Hasil transformasi menghasilkan representasi numerik dalam bentuk matriks *TF-IDF* yang digunakan dalam model *naive bayes*, hasil transformation dapat dilihat pada (Gambar 6).

	saas	saah	sall	samIn	saoukoak	abaah	abad	abadIII	abah \
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.412818
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
...
1952	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1953	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1954	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1955	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1956	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000

(Gambar 6. Data Hasil Transformation)

4.4 Data Mining

Klasifikasi dilakukan dengan algoritma *Naive Bayes Classifier*, dengan tiga skenario train-test split:

- a. Skenario 90:10: 90% data untuk pelatihan, 10% untuk pengujian.

(Tabel 1. Actual Prediction Skenario 90:10)

Prediction	Aktual Negatif	Aktual Positif
Negatif	89	16
Positif	31	60

Diperoleh 89 data komentar negatif yang berhasil diprediksi dengan benar sebagai sentimen negatif (*True Negative*), sedangkan 31 data komentar negatif yang diprediksi sebagai sentimen positif (*False Negative*). Dan terdapat 16 data komentar positif yang diprediksi sebagai sentimen negatif (*False Positive*), sedangkan 60 data komentar positif yang berhasil diprediksi dengan benar sebagai sentimen positif (*True Positive*).

- b. Skenario 80:20: 80% data untuk pelatihan, 20% untuk pengujian.

(Tabel 2. Actual Prediction Skenario 80:20)

Prediction	Aktual Negatif	Aktual Positif
Negatif	183	35
Positif	63	111

Diperoleh 183 data komentar negatif yang berhasil diprediksi dengan benar sebagai sentimen negatif (*True Negative*), sedangkan 63 data komentar negatif yang diprediksi sebagai sentimen positif (*False Negative*). Dan terdapat 35 data komentar positif yang diprediksi sebagai sentimen negatif (*False Positive*), sedangkan 111 data komentar positif yang berhasil diprediksi dengan benar sebagai sentimen positif (*True Positive*).

- c. Skenario 70:30: 70% data untuk pelatihan, 30% untuk pengujian.

(Tabel 3. Actual Prediction Skenario 70:30)

Prediction	Aktual Negatif	Aktual Positif
Negatif	278	52
Positif	97	161

Diperoleh 278 data komentar negatif yang berhasil diprediksi dengan benar sebagai sentimen negatif (*True Negative*), sedangkan 97 data komentar negatif yang diprediksi sebagai sentimen positif (*False Negative*). Dan terdapat 52 data komentar positif yang diprediksi sebagai sentimen negatif (*False Positive*), sedangkan 161 data komentar positif yang berhasil diprediksi dengan benar sebagai sentimen positif (*True Positive*).

4.5 Evaluation

Setelah melakukan klasifikasi menggunakan algoritma naive bayes classifier dengan tiga skenario pembagian data, tahap selanjutnya adalah evaluasi untuk mengukur performa model. Evaluasi dilakukan dengan menghitung beberapa metrik penting, yaitu accuracy, precision, recall, dan F1-score untuk setiap skenario. Berikut ini gambar - gambar dari hasil evaluation dari tiga skenario yang berbeda.

- a. Skenario 90:10

```

=====
                precision    recall  f1-score   support

  NEGATIVE         0.74         0.85         0.79         105
  POSITIVE         0.79         0.66         0.72          91

 accuracy                   0.76         196
 macro avg         0.77         0.75         0.75         196
    
```

(Gambar 7. Hasil Evaluation Rasio 90:10)

- b. Skenario 80:20

```

=====
                precision    recall  f1-score

  NEGATIVE         0.74         0.84         0.79
  POSITIVE         0.76         0.64         0.69

 accuracy                   0.75
 macro avg         0.75         0.74         0.74
    
```

(Gambar 8. Hasil Evaluation Rasio 80:20)

- c. Skenario 70:30

```

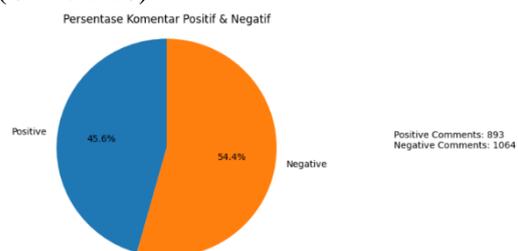
=====
                precision    recall  f1-score

  NEGATIVE         0.74         0.84         0.79
  POSITIVE         0.76         0.62         0.68

 accuracy                   0.75
 macro avg         0.75         0.73         0.74
    
```

(Gambar 7. Hasil Evaluation Rasio 90:10)

Hasil analisis sentimen ini memberikan wawasan berharga tentang persepsi publik pasca pengumuman hasil Pemilu 2024. Dominasi sentimen negatif (54,4%) mengindikasikan adanya keprihatinan masyarakat terhadap proses dan hasil pemilu. Temuan ini dapat menjadi masukan penting bagi penyelenggara pemilu serta pemerintah untuk meningkatkan transparansi dan kepercayaan publik terhadap proses demokrasi (Gambar 10).



(Gambar 10. Persentase Sentimen Negatif & Positif)

Penggunaan algoritma naive bayes classifier terbukti cukup efektif dalam mengklasifikasikan sentimen komentar, dengan performa akurasi rata – rata sebesar 75% dan terbaik untuk nilai akurasi 76%. Meskipun demikian, masih terdapat ruang untuk perbaikan, seperti mengeksplorasi algoritma klasifikasi lainnya atau menambahkan fitur yang lebih relevan untuk meningkatkan performa model.

Penelitian ini diharapkan dapat memberikan kontribusi pada pemahaman tentang dinamika opini publik di media sosial, khususnya terkait hasil pemilu Presiden 2024. Temuan ini dapat membantu pemangku kepentingan, seperti Komisi Pemilihan Umum (KPU) dan pemerintah, dalam merancang strategi komunikasi yang lebih efektif dan responsif terhadap keprihatinan masyarakat. Selain itu, hasil analisis ini juga dapat menjadi referensi bagi penelitian selanjutnya yang fokus pada analisis sentimen dalam konteks pemilu dan politik di Indonesia.

5. KESIMPULAN

Berdasarkan penelitian ini, disimpulkan bahwa sentimen negatif (54,4%) lebih dominan dibandingkan sentimen positif (45,6%) dalam diskusi publik terkait hasil Pemilu Presiden 2024 di Instagram. Implementasi algoritma naive bayes classifier pada tiga skenario pembagian data (90:10, 80:20, 70:30) menunjukkan performa rata-rata akurasi sebesar 75%, dengan skenario 90:10 memberikan hasil terbaik (accuracy 76%, precision 77%, recall 75%, F1-score 75%). Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar untuk meningkatkan akurasi, mencoba algoritma alternatif seperti K-Nearest Neighbor atau Decision Tree, melakukan pelabelan data secara manual dengan bantuan ahli, serta memanfaatkan dataset dari platform media sosial lain seperti Facebook atau Twitter untuk diversifikasi analisis.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait seperti teman – teman dan dosen yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Statista, "Share of Instagram Users by Age Indonesia," 2023. [Online]. Available: <https://www.statista.com/statistics/1078350/share-of-instagram-users-by-age-indonesia/>.
- [2] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 5, no. 3, pp. 279–285, 2019.
- [3] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 1, no. 1, pp. 42–51, 2021.
- [4] T. D. Putra, E. Utami, and M. P. Kurniawan, "Analisis Sentimen Pemilu 2024 Dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO)," *Explore*, vol. 13, no. 1, pp. 1–5, 2023.
- [5] S. A. Rismawan, "Implementasi website berita online menggunakan metode crawling data dengan bahasa pemrograman Python," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 10, no. 3, pp. 167–178, 2023.
- [6] L. Nursinggah, R. Ruuhwan, and T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis dengan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, 2024.
- [7] T. I. Hermanto and M. A. Sunandar, "Analisis Data Sebaran Penyakit Menggunakan Algoritma Density Based Spatial Clustering of Applications with Noise," *Jurnal Sains Komputer dan Teknologi Informasi*, vol. 3, no. 1, pp. 104–110, 2020.
- [8] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naive Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 145, 2019.
- [9] R. Rasenda, H. Lubis, and R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, pp. 369–376, 2020.
- [10] F. H. Rachman, "Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regression," in *2020 6th Information Technology International Seminar (ITIS)*, pp. 238–242, 2020.
- [11] H. Permana, Y. H. Chrisnanto, and H. Ashaury, "Analisis Sentimen terhadap Bakal Calon Presiden 2024 dengan Algoritma Multinomial Naive Bayes dan Oversampling SMOTE," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3257–3264, 2023.

- [12] R. Yunita and M. Kamayani, "Perbandingan Algoritma SVM dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi," **Indonesian Journal of Computer Science**, vol. 12, no. 5, 2023.
- [13] A. M. Khalimi, "Perhitungan Confusion Matrix Multi-Class Classification 3x3," **pengalaman-edukasi**, 2020. [Online]. Available: https://www.pengalaman-edukasi.com/2020/11/menghitung-confusion-matrix-3-kelas.html#google_vignette.