

# ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP SEA GAMES 2023 DI TWITTER DENGAN METODE NAÏVE BAYES

Dewi Sekar Arum<sup>1</sup>, Sufajar Butsianto<sup>2</sup>,  
Retno Astuti<sup>3</sup>

Program Studi Teknik Informatika<sup>123</sup>

Fakultas Teknik<sup>123</sup>

Univeristas Pelita Bangsa<sup>123</sup>

dewisekarar28@gmail.com<sup>1</sup>, sufajar.s@pelitabangsa.ac.id,  
retnofitriastuti13@pelitabangsa.ac.id<sup>3</sup>

**Received:** June 3 , 2023. **Revised:** July 10, 2023. **Accepted:** July 15, 2023. . **Issue Period:** Vol.7 No.3 (2023), Pp.728-738

**Abstrak:** Southeast Asian Games (SEA games) merupakan acara olahraga yang diselenggarakan di wilayah Asia Tenggara. SEA Games hanya dilaksanakan setiap dua tahun sekali yang melibatkan sebelas negara anggota *Association of South East Asian Nation* (ASEAN). Isu SEA Games di twitter banyak terjadi dikarenakan acara SEA Games 2023 banyak mengalami insiden, salah satunya bendera Indonesia yang terbalik. Tujuan dari penelitian ini untuk menganalisis sentimen masyarakat Indonesia terhadap SEA Games 2023 menggunakan metode Naïve Bayes. Data yang digunakan pada penelitian ini berasal dari data pengguna media sosial twitter yang sering menggunakan *hashtag* “SEA Games 2023” yang telah melalui text preprocessing. Hasil dari penelitian menunjukkan bahwa sebagian besar tanggapan masyarakat memiliki sentimen positif (33,4%), sentimen netral (59,1%) dan sentimen negatif (7,5%). Algoritma Naïve bayes memberikan tingkat akurasi tinggi, dengan rasio perbandingan 40:60 memiliki tingkat akurasi tertinggi (92,70%). Kata-kata yang sering muncul dalam tanggapan positif adalah "juara", "raih", "timnas", "emas", dan "medali", sedangkan dalam tanggapan negatif adalah "balik", "insiden", "kalah", dan "gagal". Penelitian ini memberikan solusi dalam memahami dan merespon tanggapan masyarakat terhadap SEA Games 2023 dengan analisis sentimen berbasis data mining di media sosial.

**Kata kunci:** Klasifikasi; Analisis Sentimen; Sea Games 2023; Naïve Bayes; Twitter

**Abstract:** The Southeast Asian Games (Sea Games) is a sports event held in the Southeast Asian region. Sea Games takes place every two years and involves eleven member countries of the Association of Southeast Asian Nations (ASEAN). The Sea Games 2023 has been a topic of discussion on Twitter due to various incidents, including the incident where the Indonesian flag was displayed incorrectly. The objective of this research is to analyze the sentiment of Indonesian society towards Sea Games 2023 using the naive Bayes method. The data used in this study is obtained from Twitter users who frequently use the hashtag "SEA Games 2023" and has undergone text preprocessing. The results of the research show that the majority of public responses have positive sentiment (33.4%), followed by neutral sentiment (59.1%) and negative sentiment (7.5%). The naive Bayes algorithm achieves a high level of accuracy, with the 40:60 data ratio showing the highest accuracy (92.70%).



DOI: 10.52362/jisamar.v7i3.1150

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

*The frequently mentioned words in positive responses are "champion," "achieve," "national team," "gold," and "medal," while in negative responses, they are "reversal," "incident," "loss," and "failure." This research provides a solution to understand and respond to public sentiment towards Sea Games 2023 through sentiment analysis based on data mining in social media.*

**Keywords:** Classification; Sentiment Analysis; Sea Games 2023; Naïve Bayes; Twitter

## I. PENDAHULUAN

Dalam era digital dan kemajuan teknologi, media sosial telah menjadi platform utama bagi pengguna untuk berbagi pendapat, opini, dan sentimen mereka terkait berbagai topik, termasuk acara olahraga seperti SEA Games. Oleh karena itu, penting untuk memahami pandangan dan opini publik melalui analisis data yang diperoleh dari media sosial.

Southeast Asian Games (SEA Games) merupakan salah satu acara olahraga terbesar di kawasan Asia Tenggara yang diadakan setiap empat tahun sekali. Selama penyelenggaraan SEA Games, media sosial terutama twitter menjadi platform yang sering digunakan oleh pengguna untuk berbagi opini dan ekspresi terkait acara tersebut. Dalam konteks SEA Games 2023, penting untuk memahami sentimen masyarakat terkait event tersebut di media sosial. Analisis sentimen masyarakat dapat memberikan wawasan tentang apakah perasaan mereka cenderung positif atau negatif terhadap SEA Games 2023. Informasi ini dapat berguna bagi penyelenggara event, pemerintah, dan pihak terkait lainnya dalam memahami persepsi masyarakat dan mengambil langkah-langkah yang tepat untuk mempromosikan dan menjalankan acara dengan sukses.

Banyaknya pengguna twitter yang menyampaikan opini-opini tersebut maka dapat dimanfaatkan untuk mencari informasi. Dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Salah satu cara yang dapat digunakan untuk menganalisis opini-opini tweet adalah dengan analisis sentimen. Metode Naïve Bayes adalah salah satu metode klasifikasi yang sering digunakan dalam analisis sentimen. Metode ini dapat memanfaatkan fitur-fitur dari data, seperti kata kunci untuk mengklasifikasikan sentimen. Namun, penggunaan metode Naïve Bayes untuk menganalisis sentimen masyarakat Indonesia terhadap SEA Games 2023 di twitter masih relatif terbatas.

Dalam penelitian [1] yang berjudul "Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization". Penelitian ini menggunakan metode Naïve Bayes dan Particle Swarm Optimization sebagai klasifikasinya, dengan tujuan untuk mengetahui hasil klasifikasi sentimen masyarakat mengenai perpanjangan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) darurat pada sosial media twitter. Hasil klasifikasi yang didapat yaitu 53,31% pengguna twitter setuju dan 46,69% pengguna twitter tidak setuju.

Dalam penelitian [2] yang berjudul "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Twitter". Dalam penelitian ini dilakukan analisis sentimen terhadap tweet yang berhubungan dengan Universitas Muhammadiyah Malang (UMM) untuk mengetahui opini masyarakat mengenai kampus. Penelitian ini menggunakan metode Naïve Bayes dan Support Vector Machine (SVM) dengan pembobotan menggunakan TF-IDF. Hasil dalam penelitian ini menunjukkan bahwa Naïve Bayes mendapatkan hasil yang lebih baik dari SVM dengan akurasi sebesar 73,65%.

Analisis sentimen atau *opinion mining* merupakan cabang ilmu dari data mining yang biasanya digunakan untuk menganalisis data tekstual berupa sebuah opini yang mengandung polaritas sehingga nantinya menghasilkan sebuah informasi yang memiliki nilai positif, negatif atau netral. [9]

Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Klasifikasi merupakan suatu teknik menentukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini mampu mengklasifikasikan data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan. [11]



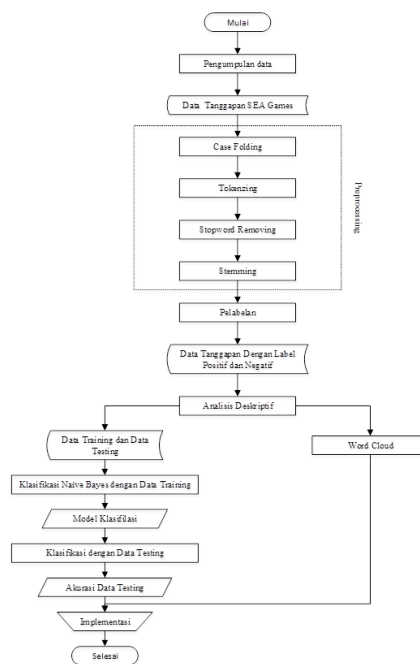
Twitter merupakan salah satu media sosial yang populer di Indonesia maupun diberbagai negara lainnya. Berdasarkan laporan *Wearesocial Hootsuite* pengguna media sosial twitter di Indonesia mencapai 52% dari jumlah pengguna internet atau mencapai 78 juta pengguna. [10]

*Southeast Asian Games* (SEA Games) merupakan salah satu *event* olahraga yang diselenggarakan di wilayah Asia Tenggara saja. Acara ini dilaksanakan setiap dua tahun sekali dan melibatkan sebelas negara anggota Association of South East Asian Nation (ASEAN). SEA Games memiliki banyak tujuan yaitu untuk mengeratkan kerjasama antar negara anggota ASEAN, menyatukan pemahaman dan mempererat hubungan antar negara anggota ASEAN di kawasan semenanjung Asia Tenggara. [14]

## II. METODE DAN MATERI

### 2.1. Tahapan Penelitian

Berikut merupakan flowchart yang menunjukan tahapan penelitian mulai dari pengumpulan hingga diperolehnya kesimpulan.



Gambar 1. *Flowchart Tahapan Penelitian*

### 2.2. Pengumpulan Data

Pengambilan data dari twitter dilakukan dengan metode scrapping data menggunakan library *snsraper*. *Twint* merupakan sebuah library untuk menscrapping data dari tweets di twitter. *Snsraper* menjadi salah satu library yang sering digunakan scraping data dari twitter karna dapat digunakan tanpa menggunakan API twitter. Pengambilan data dilakukan dari tanggal 5 juni 2023 sampai 13 juni 2023 dengan kata kunci “SEA Games”. Penelitian ini menggunakan metode dengan teknik web scrapping sedangkan software yang digunakan untuk crawling data adalah Jupyter Notebook. Melakukan crawling data pada twitter dengan library *snsraper* membutuhkan penginstallan github tujuannya untuk mengakses kedalam tweternya melalui perantara <https://github.com/JustAnotherArchivist/snsraper.git>. Setelah itu crawli-ng data dapat di lakukan dengan kode *snsraper*.

```

In [2]: import pandas as pd
import snsraper.modules.twitter as snwtwitter

In [3]: scraper = snwtwitter.TwitterSearchScraper("SEA Games lang: id since: 2023")

In [4]: tweets = scraper.get_tweets(10)
  
```

Gambar 2. kode *snsraper*

### 2.3. Preprocessing



DOI: 10.52362/jisamar.v7i3.1150

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

Data yang diperoleh dari crawling merupakan data mentah yaitu data yang masih terdapat unsur simbol, URL dan sebagainya yang tidak mempunyai arti pada kalimat tersebut. Hasil dari proses cleaning dapat dilihat pada tabel berikut:

Tabel 1. Proses *Cleaning* Data

| Sebelum Cleaning                                                                                                                                                                                                                    | Sesudah Cleaning                                                                                                                                                                                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Erick Thohir lalu memaparkan kalender Timnas Indonesia berbagai level usia yang dianggapnya berat. Mulai dari Piala AFF U-23 2023, Kualifikasi Piala Asia U-23 2024, SEA Games 2022, hingga Kualifikasi Piala Dunia 2026 Zona Asia. | Erick Thohir lalu memaparkan kalender Timnas Indonesia berbagai level usia yang dianggapnya berat Mulai dari Piala AFF U Kualifikasi Piala Asia U SEA Games hingga Kualifikasi Piala Dunia Zona Asia |
| Baca selengkapnya di : <a href="https://t.co/8w2Dxeeht2">https://t.co/8w2Dxeeht2</a><br><a href="https://t.co/ywssDDbp9G">https://t.co/ywssDDbp9G</a>                                                                               | Baca selengkapnya di                                                                                                                                                                                 |

#### 2.4. Tokenizing

Tokenizing adalah proses pemecahan teks menjadi unit-unit yang lebih kecil yang disebut token. Token dapat berupa kata-kata, frasa, tanda baca, atau elemen lainnya, tergantung pada jenis tokenizing yang digunakan. Berikut merupakan tabel dari contoh proses tokenizing.

Tabel 2 Proses Tokenizing

| Sebelum Proses Tokenizing                                                                                                                                                                              | Sesudah Proses Tokenizing                                                                                                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| garuda mendunia ketua umum pssi erick thohir menyatakan medali emas sea games kamboja menjadi momen penting bagi sepakbola indonesia memulai tradisitradiasi baru menuju kejayaan<br><br>blueprintpssi | ['garuda', 'mendunia', 'ketua', 'umum', 'pssi', 'erick', 'thohir', 'menyatakan', 'medali', 'emas', 'sea', 'games', 'kamboja', 'menjadi', 'momen', 'penting', 'bagi', 'sepakbola', 'indonesia', 'memulai', 'tradisitradiasi', 'baru', 'menuju', 'kejayaan', 'blueprintpssi'] |

#### 2.5. Stopword Removing

Stopword Removing yaitu penghapusan kata-kata yang dianggap umum dan tidak memberikan banyak informasi dalam analisis teks. Proses stopwords removing menggunakan fungsi stopwords (menghapus kata tidak penting). Pustaka seperti NLTK (Natural Language Toolkit) menyediakan koleksi stopwords dalam berbagai bahasa yang dapat digunakan untuk menghapus teks seperti kata “dan”, “ke”, “dari”, “yang”, “di”, dan lain sebagainya. Kata-kata tersebut dapat ditemukan di hampir setiap kalimat. Stopword dapat mengurangi ukuran indeks dan noise saat pemrosesan. Berikut merupakan tabel dari contoh Stopword Removing.

Tabel 3 Stopword Removing

| Sebelum Stopword Removing                                                                                                                                                                                                                                                   | Sesudah Stopword Removing                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| ['garuda', 'mendunia', 'ketua', 'umum', 'pssi', 'erick', 'thohir', 'menyatakan', 'medali', 'emas', 'sea', 'games', 'kamboja', 'menjadi', 'momen', 'penting', 'bagi', 'sepakbola', 'indonesia', 'memulai', 'tradisitradiasi', 'baru', 'menuju', 'kejayaan', 'blueprintpssi'] | garuda mendunia ketua pssi erick thohir medali emas sea games kamboja momen sepakbola indonesia tradisitradiasi kejayaan blueprintpssi |



## 2.6. Stemming

Stemming adalah proses mengubah kata-kata ke dalam bentuk dasar atau akar kata. Tujuan dilakukannya proses stemming untuk mengurangi kata-kata yang tidak memiliki akar yang sama menjadi bentuk yang lebih sederhana dan juga kata yang terbentuk memiliki representasi yang sama dengan kata yang tidak mengandung imbuhan. Berikut merupakan tabel untuk proses stemming.

Tabel 4. proses stemming

| Sebelum Stemming                                                                                                                      | Sesudah Stemming                                                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| garuda mendunia ketua pssi erick thohir medali emas sea games kamboja momen sepakbola indonesia tradisitradisi kejayaan blueprintpssi | ketua pssi erick thohir medali emas sea games kamboja momen sepakbola indonesia tradisitradisi jaya garuda dunia blueprintpssi |

## 2.7. Pembagian Data

Data set dalam penelitian ini merupakan data tanggapan mengenai SEA Games 2023 yang diambil dari platform sosial media twitter dari tanggal 1 Januari 2023 sampai dengan tanggal 13 Juni 2023. Data tanggapan yang diperoleh sebanyak 3532 tanggapan. Penentuan data training dengan perbandingan rasio untuk masing-masing data positif dan negatif sebanyak 1102 data tanggapan positif dan 247 data tanggapan negatif. Sehingga total data tanggapan yang digunakan sebanyak 1349 tanggapan. Sedangkan untuk data netral tidak digunakan karena tidak memberi informasi yang penting. Pada penelitian ini seluruh data dibagi kedalam 5 kombinasi perbandingan rasio data training dan data testing seperti pada tabel dibawah ini.

Tabel 5 Rasio Data Training dan Data Testing

| No | Jumlah Data | Rasio<br>Training : Testing (%) | Data Training | Data Testing |
|----|-------------|---------------------------------|---------------|--------------|
| 1  | 1349        | 20:80                           | 1079          | 270          |
| 2  |             | 40:60                           | 809           | 540          |
| 3  |             | 60:40                           | 539           | 810          |
| 4  |             | 80:20                           | 269           | 1080         |
| 5  |             | 90:10                           | 134           | 1215         |

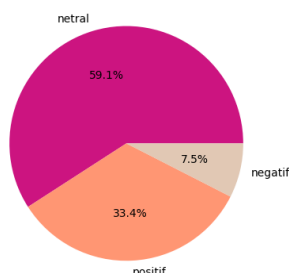
## III. PEMBAHASA DAN HASIL

### 3.1. Hasil Pengujian

Hasil analisis data tanggapan digunakan untuk mengetahui gambaran umum mengenai data tanggapan tentang SEA GAMES setelah melalui text preprocessing dan pelabelan secara otomatis oleh kamus lexicon. Data tanggapan yang diperoleh pada 05 Juni 2023 sampai 13 Juni 2023 ini dikategorikan kedalam tiga kelas kategori yaitu tanggapan positif, tanggapan negatif, dan tanggapan netral. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori ditunjukkan oleh grafik berikut.



Grafik Perbandingan Jumlah Sentimen



Gambar 3 Grafik Persentase Jumlah Tanggapan

Berdasarkan gambar grafik 4.1 diatas, diketahui bahwa dari total tanggapan yang diperoleh dari twitter yaitu 1349 tanggapan, sebanyak 1102 (33,4%) merupakan tanggapan positif yaitu tanggapan yang mendukung atau menyukai seagames sedangkan 247 (7,5%) merupakan tanggapan negatif yaitu tanggapan yang kurang menyukai suatu hal yang berhubungan dengan seagames dan untuk tanggapan netral sebanyak 2184 (59,1%) tanggapan. Dilihat dari jumlah persentase data tanggapan mengenai seagames, tanggapan positif lebih besar dibandingkan dengan tanggapan negatif dengan selisih 855 (25,9%) tanggapan.

Penentuan banyaknya data latih dan data uji ditentukan dengan perbandingan rasio yang ditunjukkan pada tabel dibawah ini.

Tabel 6 Pembagian Data Latih dan Data Uji

| Jumlah Data | Rasio Perbandingan (%) | Klasifikasi | Data Training | Data Testing |
|-------------|------------------------|-------------|---------------|--------------|
| 1349        | 20 : 80                | Positif     | 540           | 135          |
|             |                        | Negatif     | 539           | 135          |
|             |                        | Total       | 1079          | 270          |
|             | 40 : 60                | Positif     | 405           | 270          |
|             |                        | Negatif     | 404           | 270          |
|             |                        | Total       | 809           | 540          |
|             | 60 : 40                | Positif     | 270           | 220          |
|             |                        | Negatif     | 269           | 220          |
|             |                        | Total       | 539           | 440          |
|             | 80 : 20                | Positif     | 135           | 540          |
|             |                        | Negatif     | 134           | 540          |
|             |                        | Total       | 269           | 1080         |
|             | 90 : 10                | Positif     | 67            | 608          |
|             |                        | Negatif     | 67            | 607          |
|             |                        | Total       | 134           | 1215         |

Tabel diatas menunjukan pembagian data *training* dan data *testing* yang terbagi dalam 5 perbandingan rasio data dengan rasio perbandingan sebesar 20% data training : 80% data testing, 40% data training : 60% data testing, 60% data training : 40% data testing, 80% data training : 20% data testing, dan 90% data training : 10% data testing. Kemudian dilakukan *cross validation* sebanyak 10 kali.

Hasil dari average accuracy, macro average precision, dan macro average recall ditunjukkan pada tabel dibawah ini.

Tabel 7 Average Accuracy, Precision dan Recall

| Nilai                   | Naïve Bayes |
|-------------------------|-------------|
| Average Accueacy        | 0,9108622   |
| Macro Average Precision | 0,9103253   |
| Macro Average Recall    | 0,9893329   |



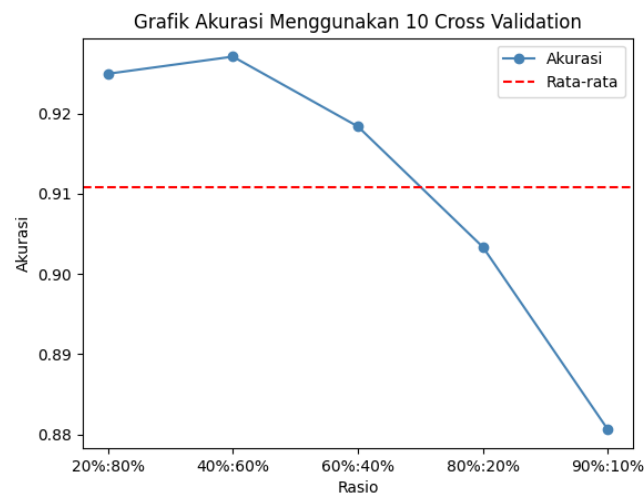
Berdasarkan tabel diatas, diketahui bahwa dengan menggunakan 5 perbandingan rasio yang berbeda-beda, algoritma *Naive Bayes* memiliki nilai *average accuracy* sebesar 0,9108622 (91,08%), nilai *macro average precision* sebesar 0,9103253 (91,03%), dan nilai *macro average recall* sebesar 0,9893329 (98,93%).

Untuk menguji performance machine learning dalam melakukan klasifikasi, maka dilakukan cross validation menggunakan 10 fold cross validation sehingga diperoleh hasil rata-rata accuracy sebesar 0,9108622, untuk lebih detailnya ditunjukkan oleh tabel dan grafik dibawah ini.

Tabel 8. Accuracy Menggunakan 10 fold Cross Validation

| Rasio Perbandingan (%) | Accuracy Naïve Bayes |
|------------------------|----------------------|
| 20 : 80                | 0,9243305            |
| 40 : 60                | 0,9270705            |
| 60 : 40                | 0,9183673            |
| 80 : 20                | 0,9033457            |
| 90 : 10                | 0,8805970            |
| Rata-rata              | 0,9108622            |

Berdasarkan tabel diatas dapat disajikan Dalam bentuk grafik berikut ini.



Gambar 4 Grafik Accuracy Menggunakan 10 Fold Cross Validation

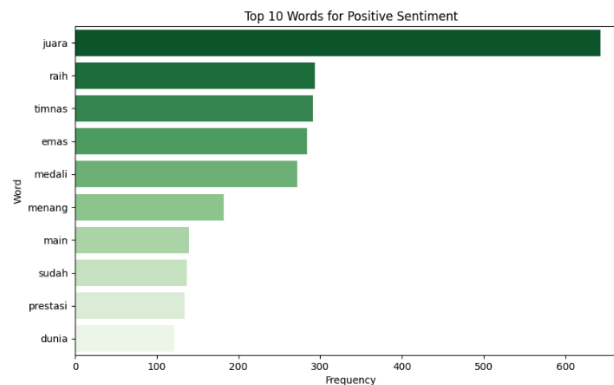
Berdasarkan gambar diatas, diketahui bahwa tingkat akurasi algoritma naïve bayes berkisar antara 88,00 % - 92,70% atau rata-rata tingkat akurasi sebesar 91,08%. Algoritma *Naive Bayes* mengalami peningkatan nilai akurasi tertinggi ketika menggunakan rasio perbandingan 40:60, kemudian mengalami penurunan akurasi terendah pada rasio 90:10.

### 3.2. Pembahasan

Pengujian kinerja machine learning difokuskan pada perbandingan rasio yang memiliki kinerja terbaik didasarkan pada tingkat akurasi yang tinggi. Perbandingan akurasi berdasarkan jumlah data uji pada algoritma naïve bayes pada masing-masing rasio berdasarkan hasil evaluasi menggunakan visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna twitter tentang seagames dengan klasifikasi ulasan positif.







Gambar 5 Frekuensi Kata Kelas Positif

pada hasil klasifikasi tanggapan positif mengenai seagames, dari jumlah ulasan positif sebanyak 1102 ulasan, diperoleh 10 kata yang paling banyak muncul diantaranya adalah kata “juara” dengan frekuensi sebanyak 643 kali, “raih” sebanyak 294 kali, “timnas” sebanyak 291 kali, “emas” sebanyak 284 kali, “medali” sebanyak 272 kali, “menang” sebanyak 182 kali, “main” sebanyak 139 kali, “sudah” sebanyak 137 kali, “prestasi” sebanyak 134 kali, “dunia” sebanyak 121 kali. Kata-kata yang muncul pada gambar diatas merupakan 10 kata yang dimiliki sentimen positif dan merupakan topik pembicaraan yang paling banyak diulas oleh pengguna twitter mengenai seagames.

Kumpulan kata-kata yang sering muncul tersebut dapat ditampilkan dalam bentuk wordcloud seperti terlihat pada gambar berikut.



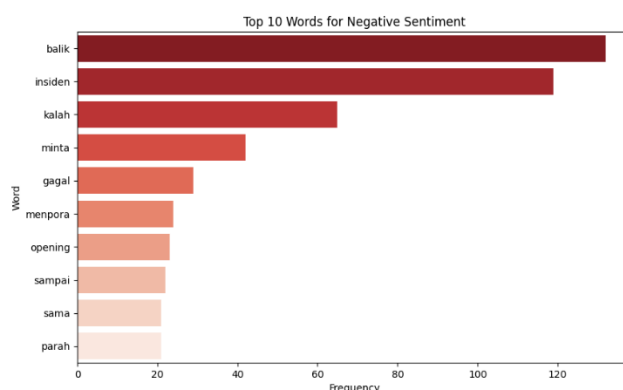
Gambar 6 Word Cloud Ulasan Positif

Visualisasi word cloud pada gambar diatas memberikan gambaran yang lebih jelas tentang topik dan kata-kata positif yang sering digunakan pengguna twitter dalam memberikan ulasan. Semakin besar ukuran kata pada word cloud menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering pengguna menggunakan kata tersebut sebagai topik pembicaraan atau penilaian positif dalam ulasan. Word cloud pada beberapa topik atau kata yang sering dibahas pengguna twitter mengenai seagames diantaranya adalah “juara”, “raih”, “timnas”, “emas”, “medali”, “menang”, “main”, “sudah”, “prestasi”, “dunia”.

Ekstraksi informasi pada ulasan negatif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan negatif yang paling sering diulas atau dibicarakan. Dari total ulasan negatif 247 ulasan, hasil ekstraksi informasi berupa ulasan negatif diidentifikasi berdasarkan frekuensi kata dalam ulasan. Berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna twitter dengan klasifikasi negatif mengenai seagames.



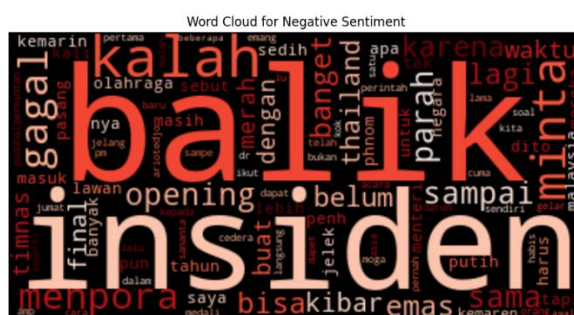




Gambar 7 Frekuensi Kata Kelas Negatif

pada hasil klasifikasi ulasan negatif pengguna twitter terhadap seagames diperoleh beberapa kata yang paling banyak muncul dengan topik yang dianggap relevan sebagai sentimen negatif diantaranya adalah kata “balik” dengan frekuensi sebanyak 132 kali, “insiden” sebanyak 119 kali, “kalah” sebanyak 65 kali, “minta” sebanyak 42 kali, “gagal” sebanyak 29 kali, “menpora” sebanyak 24 kali, “opening” sebanyak 23 kali, “sampai” sebanyak 22 kali, “sama” sebanyak 21 kali, “parah” sebanyak 21 kali. Kata-kata yang muncul pada gambar tersebut merupakan kata yang dimiliki sentimen negatif dan merupakan kata atau topik pembicaraan yang paling banyak diulas oleh pengguna twitter mengenai seagames.

Kumpulan kata-kata yang sering muncul tersebut ditampilkan dalam bentuk word cloud seperti terlihat pada gambar berikut.



Gambar 8 Word Cloud Ulasan Positif

Visualisasi word cloud pada gambar tersebut memberikan gambaran yang lebih jelas tentang topik dan kata-kata negatif yang sering digunakan pengguna twitter dalam memberikan tanggapan mengenai seagames. Beberapa topik yang sering dibahas pengguna twuiter tentang seagames diantaranya adalah kata “balik”, “insiden”, “kalah”, “minta”, “gagal”, “menpora”, “opening”, “sampai”, “sama”, “parah”.

#### IV. KESIMPULAN

Dalam analisis sentimen, ditemukan bahwa sebagian besar tanggapan masyarakat terhadap sea games 2023 di twitter memiliki sentimen positif sebesar (33,4%), diikuti oleh sentimen netral sebesar (59,1%), dan sentimen negatif sebesar (7,5%). untuk klasifikasi tanggapan terhadap SEA Games 2023 menunjukan tingkat akurasi yang tinggi. Rasio perbandingan data training dan data testing berpengaruh terhadap kinerja algoritma, dengan berbanding 40:60 menunjukan tingkat akurasi (92,70%). 3. Dengan menggunakan beberapa perbandingan data latih dan data uji sebesar 20%:80%, 40%:60%, 60%:40%, 80%:20%, dan 90%:10% diperoleh hasil klasifikasi sentimen negatif dan positif dengan metode naïve bayes dan menggunakan 10-fold cross validation diperoleh nilai rata-rata akurasi sebesar 91,09%.



DOI: 10.52362/jisamar.v7i3.1150

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

## REFERENASI

- [1] S. D. Pramukti, A. Nugroho, dan A. S. Sunge, “Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization,” *Techno.Com*, vol. 21, no. 1, hal. 61–74, 2022, doi: 10.33633/tc.v21i1.5332.
- [2] M. I. Fikri, T. S. Sabrila, dan Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika J.*, vol. 10, no. 02, hal. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [3] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, dan R. Damaiyanti, “Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification,” *Sistemasi*, vol. 10, no. 1, hal. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [4] D. Duei Putri, G. F. Nama, dan W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, hal. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [5] M. I. Petiwi, A. Triayudi, dan I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 6, no. 1, hal. 542, 2022, doi: 10.30865/mib.v6i1.3530.
- [6] A. R. Isnain, A. I. Sakti, D. Alita, dan N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, hal. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [7] D. A. Agustina, S. Subanti, dan E. Zukhronah, “Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine,” *Indones. J. Appl. Stat.*, vol. 3, no. 2, hal. 109, 2021, doi: 10.13057/ijas.v3i2.44337.
- [8] H. D. Wijaya dan S. Dwiasnati, “Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat,” *J. Inform.*, vol. 7, no. 1, hal. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [9] B. Laurensz dan Eko Sedyono, “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, hal. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [10] N. Fitriyah, B. Warsito, dan D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm,” *J. Gaussian*, vol. 9, no. 3, hal. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [11] I. Romli dan A. T. Zy, “Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, hal. 694–702, 2020.
- [12] D. P. Utomo dan M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, hal. 437, 2020, doi: 10.30865/mib.v4i2.2080.



DOI: 10.52362/jisamar.v7i3.1150

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

- [13] L. O. Sihombing, H. Hannie, dan B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, hal. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [14] M. Riefky dan W. Pramesti, “Sentiment Analysis of Southeast Asian Games (SEA Games) in Philippines 2019 Based on Opinion of Internet User of Social Media Twitter with K-Nearest Neighbor and Support Vector Machine,” *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, hal. 26–41, 2020, doi: 10.20956/jmsk.v17i1.9947.
- [15] A. R. Rizqullah, A. Wedhasmara, R. I. Heroza, A. Putra, F. Fathoni, dan P. Putra, “Analisis Masalah Pada Data Review Aplikasi Terhadap Layanan E-Commerce Menggunakan Metode Text Classification,” *J. Tekno Kompak*, vol. 16, no. 1, hal. 186, 2022, doi: 10.33365/jtk.v16i1.1448.
- [16] D. Normawati dan S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 697–711, 2021, [Daring]. Tersedia pada: <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- [17] B. Warna *et al.*, “SmartAI,” vol. 1, no. 4, hal. 201–206, 2022.
- [18] I. Technology, “Journal of Vocational Education and Information Technology,” vol. 3, no. 2, hal. 65–70, 2023.
- [19] S. R. Pratama dan A. H. Mirza, “Penerapan Data Mining Untuk Memprediksi Tingkat Inflasi Menggunakan Metode Regresi Linier Berganda Pada BPS,” *Bina Darma Conf. Comput. Sci.*, hal. 245–255, 2021.
- [20] A. Kreteria, P. Dengan, dan T. Semantik, “Implementasi aplikasi jupyter notebook sebagai analisis kreteria plagiasi dengan teknik semantik,” vol. 8, no. 2, hal. 627–637, 2023.

