

Pemanfaatan *Word Cloud* Pada Analisis Sentimen Dalam Menggali Persepsi Publik

Paulina Agusia¹, Merry Uli Artha Manurung², Vincent Calista³, Viny Christanti Mawardi⁴

Program Studi Teknik Informatika
Universitas Tarumanagara
Jakarta, Indonesia

e-mail: ¹paulina.535220048@stu.untar.ac.id, ²merry.535220263@stu.untar.ac.id,
³vincent.535220075@stu.untar.ac.id, ⁴viny@fti.untar.ac.id

Correspondence : e-mail: viny@fti.untar.ac.id

Diajukan: 30 Juni 2024; Direvisi: 09 Agustus; Diterima: 28 Agustus 2024

Abstrak

Penggunaan *Word Cloud* dalam analisis sentimen telah terbukti sebagai metode yang efektif untuk menggali persepsi publik. *Word Cloud*, yang merupakan representasi visual dari frekuensi kata dalam sebuah teks, memungkinkan para peneliti untuk dengan cepat mengidentifikasi tema dan topik utama dalam data teks yang besar. Dalam konteks analisis sentimen, *Word Cloud* membantu dalam memahami penyebaran dan frekuensi kata-kata yang bermuatan emosional positif atau negatif dalam percakapan publik. Studi ini meneliti penerapan *Word Cloud* dalam analisis sentimen terhadap data dari media sosial dan platform aplikasi X untuk mengungkap persepsi dan opini masyarakat tentang berbagai isu. Hasil penelitian menunjukkan bahwa *word cloud* tidak hanya mempermudah visualisasi data, tetapi juga meningkatkan akurasi dan efisiensi dalam mendeteksi tren sentimen. Oleh karena itu, *word cloud* menjadi alat yang sangat berguna dalam analisis sentimen untuk mengidentifikasi dan memahami persepsi publik secara mendalam.

Kata kunci: Analisis Sentimen, *Word Cloud*, Analisis Sentimen Twitter, Naive Bayes, TextBlob.

Abstract

The utilization of *Word Cloud* in sentiment analysis demonstrated effectiveness in uncovering public perceptions. A *Word Cloud*, being a visual representation of word frequency within a text, enabled researchers to swiftly pinpoint the main themes and topics within large text datasets. In sentiment analysis, *Word Clouds* assisted in understanding how frequently positive or negative emotionally charged words appeared in public discussions. This study investigated the application of *Word Clouds* for sentiment analysis using data from social media and the X application platform to uncover public opinions and perceptions on various issues. The results revealed that *Word Clouds* not only simplified data visualization but also improved the accuracy and efficiency of identifying sentiment trends. Consequently, *Word Clouds* became a valuable tool in sentiment analysis for deeply understanding and identifying public perceptions.

Keywords: Sentiment Analysis, *Word Cloud*, Twitter Sentiment Analysis, Naive Bayes, TextBlob.

1. Pendahuluan

Dalam era digital saat ini, jumlah data tekstual yang dihasilkan oleh pengguna di media sosial dan platform online lainnya terus meningkat secara signifikan. Data ini berisi beragam opini, sentimen, dan persepsi publik mengenai topik tertentu dan sangat berharga bagi para peneliti, pemasar, dan pembuat kebijakan. Salah satu tantangan terbesar saat menganalisis data ini adalah mengidentifikasi dan memahami sentimen dalam data secara efisien dan efektif. *Word Cloud* populer sebagai alat visualisasi yang efektif untuk analisis teks. *Word Cloud* menampilkan frekuensi kata dalam format visual yang mudah dipahami, memungkinkan peneliti dengan cepat mengidentifikasi kata kunci utama dan tema utama dalam korpus teks [1]-[2].

Word Cloud dapat membantu mengidentifikasi pola dan tren dalam data teks berskala besar. Selain itu, *Word Cloud* dapat digunakan untuk mewakili sebaran sentimen dalam sebuah teks dengan menyorot kata-kata yang berkonotasi positif atau negatif [3].

Masalah utama dalam analisis sentimen adalah menyajikan hasil analisis secara intuitif dan mudah dipahami. Metode tradisional seperti analisis statistik dan pemrosesan bahasa alami (NLP) seringkali memerlukan pemahaman teknis yang mendalam dan sulit diakses oleh banyak pengguna [4].

Hal ini menyebabkan diperlukannya suatu metode untuk mempermudah proses analisis dan visualisasi data sentimen. Penelitian ini mengusulkan penggunaan *Word Cloud* sebagai pendekatan inovatif dalam analisis sentimen untuk mempelajari persepsi masyarakat. Metode ini menggabungkan teknik NLP untuk klasifikasi sentimen dengan visualisasi *Word Cloud* dan menyajikan hasil analisisnya [5]. Penelitian ini menggunakan data yang diperoleh dari media sosial dan platform aplikasi X serta mencakup berbagai opini masyarakat terhadap isu-isu terkini. Kontribusi baru dari penelitian ini adalah pengembangan pendekatan analisis sentimen yang lebih intuitif dan efisien menggunakan *Word Cloud* [6]-[7].

2. Metode Penelitian

Metode penelitian mencakup beberapa langkah diantaranya pengumpulan data, *preprocessing data*, klasifikasi, dan evaluasi. Berikut adalah tahapan rancangan penelitian:

1. Metode *Crawling*

Pengumpulan informasi dengan menghimpun data secara otomatis menggunakan program khusus yang disebut dengan *crawler* yaitu alat untuk melakukan *crawling* dengan cara menjelajahi web untuk mengunduh dan menyimpan informasi dari halaman web [8].

2. Pembersihan Data (*Data Cleaning*)

Data cleaning adalah salah satu proses penting dalam analisis data yang berfungsi untuk mengidentifikasi, memperbaiki, serta menghapus dataset yang tidak konsisten. *Data cleaning* juga bertugas untuk memastikan bahwa data yang digunakan dalam analisis adalah akurat, lengkap, dan siap digunakan [9].

3. Pra-pemrosesan Data (*Data Preprocessing*)

Preprocessing data adalah serangkaian langkah yang diterapkan pada dataset sebelum data tersebut digunakan untuk analisis. *Preprocessing data* berfungsi untuk mempersiapkan data sehingga menjadi lebih mudah dipahami oleh komputer serta lebih cocok untuk analisis atau pengolahan lebih lanjut [10]-[11]. Berikut adalah proses dibalik *preprocessing data*:

1) Normalisasi (*Normalization*)

Normalisasi adalah proses mengganti kata-kata tertentu dengan sinonim atau bentuk yang lebih umum [12].

2) *Stopwords*

Stopwords adalah kata-kata umum yang biasanya hanya memberikan sedikit kontribusi terhadap makna teks (misalnya, "dan", "dalam", "yang").

3) Tokenisasi (*Tokenization*)

Tokenisasi merupakan sebuah proses pengelompokan suatu rangkaian karakter berdasarkan karakter spasi, dan pada waktu yang bersamaan dilakukan juga proses penghapusan karakter tertentu, seperti tanda baca.

4) *Stemming*

Kata-kata yang sudah diubah menjadi huruf kecil perlu dilakukan pengecekan. *Stemming* digunakan untuk membuat kata-kata yang seragam sehingga mengurangi daftar kata yang ada pada data latih.

5) Proses Pemberian Label (*Labeling*)

Terdapat 2 (dua) jenis *Labeling* untuk mengidentifikasi suatu variabel atau bagian dari data atau berkas. yaitu pemberian label kepada token dengan kata penguat (*exaggeration*) dan kata negasi (*negation*).

Penelitian ini menggunakan *Word Cloud* untuk merepresentasi visual teks, di mana kata-kata yang paling sering muncul diberi penekanan visual yang lebih besar. Proses pembuatan *Word Cloud* melibatkan pengumpulan teks dan menghitung frekuensi setiap kata, kemudian ditampilkan dengan ukuran yang sebanding dengan frekuensinya. *Word Cloud* memungkinkan kita untuk melihat kata-kata yang paling sering muncul dalam setiap kategori visual sentimen, di mana ukuran kata dalam *Word Cloud* menunjukkan seberapa sering kata tersebut muncul dalam teks.

3. Hasil dan Pembahasan

Dataset pada penelitian ini diambil dari file CSV yang dikumpulkan dari aplikasi X dengan melakukan *crawling*. Data yang diambil berupa *tweet* dalam bahasa Indonesia dengan kata kunci “Gibran” dengan jumlah data sebanyak 367 data.

3.1. Tahapan *Processing*

Tahap *processing* dalam analisis sentimen meliputi pengumpulan data teks, *preprocessing* untuk membersihkan dan menormalisasikan teks dengan menggunakan teknik seperti *Bag of Words* atau TF-IDF, lalu dianalisis untuk mengidentifikasi kata kunci yang mempengaruhi sentimen dan divisualisasikan dalam bentuk *Word Cloud*.

3.1.1. Normalisasi

Normalisasi dilakukan dengan tujuan untuk mencegah terjadinya kesalahan ejaan dalam kata agar lebih mudah diproses dan di analisis. Berikut merupakan data sebelum dan sesudah dinormalisasi:

```
Gibran berhasil bikin Solo eksis, Cak imin mun..
↓
gibran berhasil bikin solo eksis cak imin mun..
```

Gambar 1. Normalisasi

3.1.2. *Stopword*

Stopword dilakukan guna untuk menghapus kata penghubung yang tertulis pada komentar. Berikut merupakan data sebelum dan sesudah dilakukan *Stopword*:

```
solo yang sukses di bawah gibran adalah bukti
↓
solo sukses bawah gibran bukti perubahan
```

Gambar 2. *Stopword*

3.1.3. Tokenisasi

Tujuan tokenisasi dilakukan untuk mengurai teks menjadi kata individual, yang kemudian dapat diproses lebih lanjut untuk memahami makna dan sentimen yang terkandung pada sebuah kata. Berikut adalah gambar sebelum dan sesudah dilakukan tokenisasi:

```
solo sukses bawah gibran bukti perubahan
↓
[solo, sukses, bawah, gibran, bukti, perubahan]
```

Gambar 3. Tokenisasi

3.1.4. *Stemming*

Stemming berfungsi untuk menyederhanakan kata-kata menjadi bentuk dasarnya, membantu mengurangi kompleksitas teks, meningkatkan akurasi pencarian informasi, mengoptimalkan algoritma pembelajaran sebuah mesin, dan memperbaiki analisis teks [13].

```
gibran berhasil bikin solo eksis,
↓
gibran hasil bikin solo eks cak imin
```

Gambar 4. *Stemming*

3.2. Klasifikasi

3.2.1. *Naive Bayes*

Pada Tabel 1, menunjukkan bahwa *Naive Bayes* digunakan untuk mengklasifikasikan teks berdasarkan sentimen yang diprediksikan dalam teks tersebut serta untuk mendapatkan secara otomatis mengklasifikasi *tweet* sebagai sentimen yang positif, netral, atau negatif berdasarkan data latih sebanyak 93 data *tweet*.

Tabel 1. Klasifikasi *Naive Bayes*.

| full_text | tweet_english | Klasifikasi | klasifikasi_bayes |
|---|--|-------------|-------------------|
| gibran bikin solo eks cak imin mungkin cari.. | gibran make solo ex cak imin maybe.. | Positif | Positif |
| dibilangin benar gibran samsul android bukan.. | said to be true gibran samsul android is.. | Negatif | Negatif |
| gibran hasil bikin solo eks cak imin mungkin.. | gibran results make solo cak imin may.. | Netral | Positif |
| ijin tamu | guest permissions | Netral | Netral |
| gibran bikin solo jadi pusat perhati cak imin.. | gibran make solo the center of attention.. | Netral | Positif |

Tabel 1 menunjukkan metode ini merupakan sebuah algoritma klasifikasi yang didasarkan pada Teorema *Bayes*, memanfaatkan probabilitas untuk membuat keputusan yang prediktif [5]. Berikut adalah rumus dasar *Naive Bayes*:

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)} \quad (1)$$

Pada rumus Persamaan 1, hipotesis (h) mengacu pada data yang diasosiasikan dengan suatu kelas tertentu. Data (D) adalah kumpulan data yang belum memiliki kelas. Probabilitas hipotesis ($P(h)$) atau *Prior Probability* adalah probabilitas awal dari hipotesis sebelum memperhitungkan data. Probabilitas $P(D)$ adalah probabilitas data tersebut terjadi. *Posterior Probability* ($P(h|D)$) adalah probabilitas hipotesis setelah mempertimbangkan data, sedangkan $P(D|h)$ adalah probabilitas data berdasarkan kondisi pada hipotesis tersebut. Dalam konteks ini, *Naive Bayes* menghitung probabilitas setiap kelas berdasarkan fitur dalam data dan kemudian memilih kelas dengan probabilitas tertinggi sebagai prediksi sentimen dari sebuah teks [10]-[11].

Dalam analisis sentimen, terdapat proses umum yang melibatkan proses pengumpulan data yang telah dikategorikan dalam sentimen positif, netral, atau negatif, lalu *preprocessing* untuk membersihkan data yang mencakup tokenisasi, penghapusan *stopword* dan *stemming*. Selain itu juga dilakukan pelatihan model menggunakan data latih untuk menghitung *prior probability* dan *conditional probability* serta mengklasifikasi teks baru berdasarkan probabilitas yang telah dipelajari [3]-[5].

3.2.2. TextBlob

TextBlob akan mengembalikan nilai polaritas yang dimana jika nilai polaritas sama dengan 0 (nol) artinya bersifat netral, polaritas lebih dari 0 bersifat positif, dan polaritas dibawah 0 bersifat negatif seperti pada Tabel 2.

Tabel 2. Klasifikasi *TextBlob*.

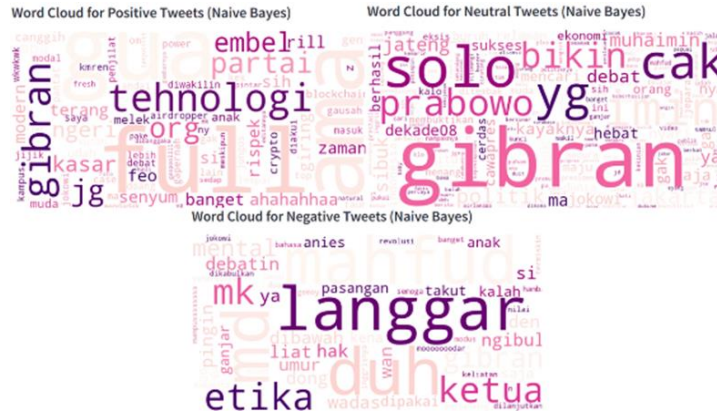
| Tweet | Sentimen | Polaritas | Kata Positif | Kata Netral |
|--|----------|-----------|--------------|-------------|
| gibran make solo ex cak imin maybe find a political stage | Positif | 0.29 | 1 | 11 |
| said to be true gibran samsul android is not a human being, | Negatif | -0.42 | 2 | 42 |
| gibran results make solo ex cak imin maybe teach how to | Netral | 0 | 0 | 16 |
| guest permissions | Netral | 0 | 0 | 2 |
| gibran makes solo the center of attention of cak imin seems to | Netral | 0 | 1 | 20 |

3.3. Visualisasi *Word Cloud*

Word Cloud adalah visualisasi kata-kata dari sebuah teks, di mana kata-kata yang paling sering muncul ditampilkan lebih besar. *Word cloud* membantu secara visual menggambarkan frekuensi dan signifikansi relatif dari kata-kata dalam teks, memudahkan pemahaman tema atau fokus utama dengan cepat.

3.3.1. *Word Cloud Naive Bayes*

Pada Gambar 5 menunjukkan Visualisasi *Word Cloud* dengan metode *Naive Bayes* bertujuan untuk menyoroti kata-kata penting yang berkontribusi pada klasifikasi sentimen positif, negatif, dan netral. Kata-kata yang paling berpengaruh pada klasifikasi sentimen ditampilkan pada *Word Cloud*, dengan ukuran kata yang mencerminkan seberapa signifikan kata tersebut dalam sebuah klasifikasi sentimen. Hal ini mempermudah untuk memahami elemen teks yang berperan besar dalam menganalisis sentimen dan memberikan wawasan tentang bagaimana metode *Naive Bayes* membuat prediksi [2].



Gambar 5. Visualisasi Word Cloud Naive Bayes

3.3.2. Word Cloud TextBlob

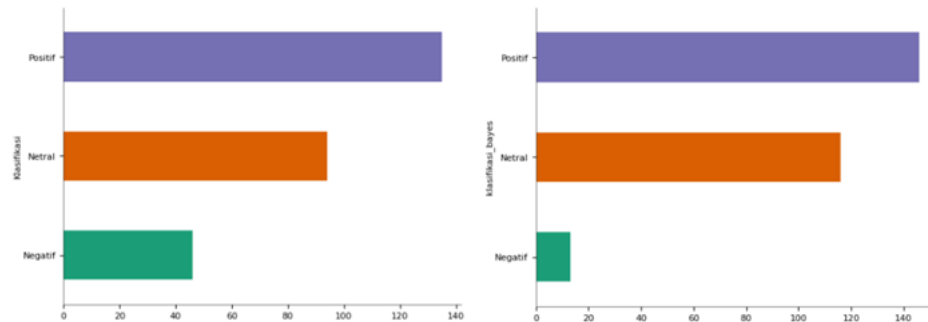
Visualisasi Word Cloud dengan menggunakan metode TextBlob seperti pada Gambar 6 untuk memunculkan teks yang paling sering muncul pada data yang di analisis dan dibuat per-sentimen (positif, negatif, dan netral).



Gambar 6. Visualisasi Word Cloud TextBlob

3.4. Distribusi Kategori (Categoricals Distributions)

Terdapat 367 data tweet dari data yang dikumpulkan secara keseluruhan dimana digunakan 2 metode yaitu Naive Bayes dan TextBlob yang menghasilkan sebagai berikut:



Gambar 7. Categoricals Distributions

Grafik pada Gambar 9 menunjukkan bahwa klasifikasi dengan Naive Bayes menghasilkan 146 tweet positif, 116 tweet netral, dan 13 tweet negatif. Sedangkan TextBlob menghasilkan 135 tweet positif, 94 tweet netral, dan 46 tweet negatif. Pada hasil yang diberikan membuktikan bahwa Naive Bayes lebih mampu dalam menyesuaikan model dengan data latih, membuatnya menjadi lebih efektif dalam menangkap pola yang kompleks. Sementara TextBlob bergantung pada kata kunci atau pola tertentu dalam sebuah teks.

4. Kesimpulan

Penelitian ini menggunakan *Word Cloud* untuk analisis sentimen dari data media sosial dengan metode *Naive Bayes* dan *TextBlob*. Hasilnya menunjukkan *Naive Bayes* lebih akurat, sementara *Word Cloud* membantu visualisasi kata-kata kunci dalam data teks untuk memperjelas sentimen. Pendekatan ini menggabungkan NLP dan visualisasi untuk menyederhanakan analisis bagi audiens non-teknis. Penelitian juga mengevaluasi kekuatan dan keterbatasan *Word Cloud* serta memberikan rekomendasi untuk penerapan di masa depan.

Dengan menggunakan metode *Naive Bayes* dan *TextBlob* memberikan kontribusi yang berharga dalam memahami persepsi masyarakat. Untuk masa yang akan datang, arah penelitian terkait dapat meliputi pengembangan teknik klasifikasi yang lebih canggih, integrasi dengan machine learning yang lebih maju, dan eksplorasi visualisasi data yang lebih interaktif. Selain itu, fokus pada pengembangan algoritma untuk mengatasi isu-isu kompleks dalam analisis sentimen seperti sarcasm dan konteks budaya juga bisa menjadi arah penelitian yang menarik.

Daftar Pustaka

- [1] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, 2016, doi: 10.1186/s41044-016-0014-0.
- [2] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," *Kurdistan J. Appl. Res.*, pp. 54–65, 2020, doi: 10.24017/covid.8.
- [3] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma *Naive Bayes*," *INSOLOGI J. Sains dan Teknol.*, vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.
- [4] I. Iin, R. Supriatna, M. Mulyawan, and D. Rohman, "Penerapan Natural Language Processing Dalam Analisis Sentimen Cawapres 2024 Menggunakan Algoritma *Naive Bayes*," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 1109–1115, 2024, doi: 10.36040/jati.v8i1.8572.
- [5] Dedi Darwis, Nery Siskawati, and Zaenal Abidin, "Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2020.
- [6] H. A. Shehu *et al.*, "Deep Sentiment Analysis: A Case Study on Stemmed Turkish Twitter Data," *IEEE Access*, vol. 9, pp. 56836–56854, 2021, doi: 10.1109/ACCESS.2021.3071393.
- [7] I. G. S. Mas Diyasa, N. M. I. Marini Mandenni, M. I. Fachrurrozi, S. I. Pradika, K. R. Nur Manab, and N. R. Sasmita, "Twitter Sentiment Analysis as an Evaluation and Service Base On Python *Textblob*," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1125/1/012034.
- [8] K. S. S. Varma, P. Sathineni, and R. Mamidi, "Sentiment Analysis in Code-Mixed Telugu-English Text with Unsupervised Data Normalization," *Int. Conf. Recent Adv. Nat. Lang. Process. RANLP*, no. September, pp. 753–760, 2021, doi: 10.26615/978-954-452-072-4_086.
- [9] J. Wang, X. Wang, Y. Yang, H. Zhang, and B. Fang, "A review of data cleaning methods for web information system," *Comput. Mater. Contin.*, vol. 62, no. 3, pp. 1053–1075, 2020, doi: 10.32604/cmc.2020.08675.
- [10] Ericha Apriliyani and Y. Salim, "Analisis performa metode klasifikasi *Naive Bayes* Classifier pada Unbalanced Dataset," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 47–54, 2022, doi: 10.56705/ijodas.v3i2.45.
- [11] D. Duei Putri, G. F. Nama, and 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, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [12] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Front. Energy Res.*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/fenrg.2021.652801.
- [13] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling (Requirements Identification for NTB People in pandemic covid-19 at Social Media Using Crawling Method)," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 244–250, 2021.