

Penerapan Algoritma Naïve Bayes Classifier dalam Analisis Sentimen Masyarakat Terhadap Hasil Pemilu Presiden RI 2024

Yenni Yunnita¹, Rasyidah^{2,*}, Fazrol Rozi³

^aJurusan Teknologi Informasi, Politeknik Negeri Padang, Padang, 25176, Indonesia

* rasyidah@pnp.ac.id

| ARTICLE INFO | ABSTRACT |
|---|--|
| Article history Received 04/11/2024 Revised 14/12/2024 Accepted 09/01/2025 | <p>Sentiment analysis is the process of analyzing text to determine whether a sentence has a positive, negative or neutral message. To determine positive, negative and neutral sentiment we need to do labeling. In this research, labeling uses 4 libraries, namely, Indonesian lexicon, textblob, nltk, and vader sentiment. The results of the labeling of each library of the 6 election candidates can be concluded that the highest percentage of positive sentiment is candidate Ganjar Pranowo, namely 63.78%, the second is Mahfud MD, namely 58.91%, the third is Anies, namely 48.92%, the fourth is Muhaimin, 46.35%, the fifth is Prabowo, namely 45.84%, the sixth gibran is 43.74%. The most negative sentiment was Anies Baswedan, namely 28.68%, Muhaimin Iskandar, namely 28.65%, Prabowo Subianto, namely 26.95%, Giban Rakabuming Raka, namely 26.66%, Ganjar Pranowo, namely 23.73%, and Mahfud, namely 22.40%. So it can be concluded that public sentiment regarding the 2024 election has a lot of positive sentiment towards candidate Ganjar Pranowo and negative towards Anies Baswedan. The highest accuracy results using the Naïve Bayes algorithm were using the Indonesian lexicon library at 81.54%, second using the Vader Sentiment library at 80.91%, third using the NLTK library at 80.08%, and fourth using the Textblob library at 73.54%.</p> <p>Keywords: Sentiment analysis, Labeling, Naïve Bayes.</p> |
| | ABSTRAK <p>Analisis sentimen merupakan proses menganalisis teks untuk menentukan apakah sebuah kalimat tersebut mempunyai pesan positif, negatif, atau netral. Untuk menentukan sentiment positif, negatif, dan netral kita perlu melakukan pelabelan. Pada penelitian ini pelabelan menggunakan 4 library yaitu, lexicon bahasa indonesia, textblob, nltk, dan vader sentiment. Hasil pelabelan dari masing-masing library dari 6 kandidat pemilu ini dapat disimpulkan bahwa persentase sentiment positif paling banyak yaitu kandidat ganjar pranowo yaitu sebesar 63.78%, kedua mahfud md yaitu 58.91%, ketiga anies yaitu 48.92%, keempat muhaimin 46.35%, kelima prabowo yaitu 45.84%, keenam gibran yaitu 43.74%. Sentimen negatif paling banyak yaitu anies baswedan yaitu 28.68%, muhaimin iskandar yaitu 28.65%, prabowo subianto yaitu 26.95%, giban rakabuming raka yaitu 26.66%, ganjar pranowo yaitu 23.73%, dan mahfud yaitu 22.40%. Jadi disimpulkan bahwa sentiment masyarakat terkait pemilu 2024 banyak sentiment positif kepada kandidat ganjar pranowo dan negative kepada anies baswedan. Hasil akurasi paling tinggi menggunakan algoritam naïve bayes yaitu menggunakan library lexicon bahasa indonesia sebesar 81.54%, kedua menggunakan library vader sentiment sebesar 80.91%, ketiga menggunakan library nltk sebesar 80.08%, dan keempat menggunakan library textblob sebesar 73.54%.</p> <p>Kata kunci: Analisis sentimen, Pelabelan, Naïve Bayes.</p> <p>This is an open access article under the CC-BY-SA license.</p>  |

1. Pendahuluan

Pemilihan Umum (Pemilu) merupakan mekanisme yang digunakan untuk mewujudkan kedaulatan rakyat dan menghasilkan pemerintahan negara yang demokratis, sesuai dengan Pancasila dan UUD Negara RI Tahun 1945. Pemilu bertujuan untuk memilih Presiden dan Wakil Presiden, Anggota DPR, DPD, DPRD, serta kepala daerah dan wakil kepala daerah yang mampu

mencerminkan nilai-nilai demokrasi dan mampu memperjuangkan aspirasi rakyat sesuai dengan perkembangan kehidupan berbangsa dan bernegara. Pelaksanaan Pemilu dilakukan oleh penyelenggara pemilu yang memiliki integritas, profesionalitas, dan akuntabilitas. Pemilu dijalankan dengan kualitas yang baik, sistematis, sah secara hukum, dan akuntabel, dengan melibatkan partisipasi masyarakat secara luas. Oleh karena itu, penyelenggara pemilu, aparat pemerintah, peserta pemilu, pengawas pemilu, pemantau pemilu, pemilih, dan semua pihak terkait diharapkan bersikap dan bertindak jujur sesuai dengan peraturan perundang-undangan yang berlaku.

Dalam pelaksanaannya, seringkali muncul isu-isu terkait kecurangan, polarisasi politik, dan pendapat publik yang beragam terhadap calon atau partai politik tertentu. Hal ini menunjukkan bahwa proses pemilu tidak hanya menjadi ajang kompetisi politik, tetapi juga menjadi momen krusial dalam menentukan kepercayaan masyarakat terhadap sistem demokrasi itu sendiri. Oleh karena itu, pemilu harus dilaksanakan dengan kualitas yang lebih baik untuk menjamin kompetisi yang sehat, partisipatif, dengan tingkat keterwakilan yang lebih tinggi, serta memiliki mekanisme pertanggungjawaban yang jelas. Pemilu yang bersih dan adil merupakan fondasi penting untuk memastikan stabilitas politik dan sosial dalam suatu negara[1].

Pada Pemilu 2024 di Indonesia, muncul berbagai isu yang menimbulkan kekhawatiran terkait keabsahan dan integritas proses pemilu. Salah satu isu yang mencuat adalah adanya surat suara yang sudah tercoblos sebelum dilakukan pemungutan suara. Kejadian ini menimbulkan kecurigaan akan adanya praktik kecurangan yang terorganisir untuk memanipulasi hasil pemilu. Selain itu, ditemukan juga kasus surat suara yang tertukar, kurang, hilang, dan rusak. Hal ini menunjukkan adanya kelemahan dalam manajemen logistik dan distribusi surat suara yang dapat berdampak negatif terhadap kepercayaan masyarakat terhadap proses pemilu[2].

Masalah lain yang tidak kalah serius adalah kotak suara yang tidak bersegel. Kotak suara yang tidak bersegel membuka peluang untuk manipulasi dan penyalahgunaan suara, yang dapat merusak integritas hasil pemilu. Keamanan kotak suara sangat krusial untuk menjaga keabsahan suara yang telah diberikan. Selain itu, penggunaan Sistem Informasi Rekapitulasi Pemilihan Umum (SIREKAP) yang belum siap juga menimbulkan kejanggalan. Terdapat perbedaan antara jumlah perolehan suara partai dengan jumlah akumulasi perolehan suara yang didapatkan, serta perubahan drastis angka perolehan suara setelah dipindai ke aplikasi SIREKAP. Kejadian-kejadian ini memperlihatkan bahwa sistem teknologi yang digunakan dalam pemilu masih perlu perbaikan dan penyempurnaan[3].

Akan tetapi, meskipun berbagai masalah tersebut muncul, penting untuk diingat bahwa teknologi dan inovasi dalam proses pemilu juga memiliki potensi besar untuk meningkatkan transparansi dan akuntabilitas. Penggunaan teknik *text mining* dan algoritma Naive Bayes

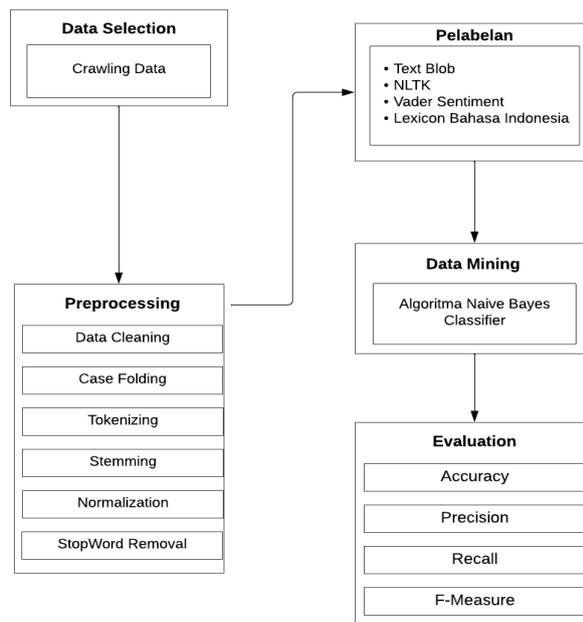
Classifier dalam analisis sentimen memungkinkan untuk memahami opini publik secara lebih mendalam dan real-time. Melalui analisis sentimen, peneliti dapat mengklasifikasikan tweet menjadi sentimen positif, negatif, atau netral, yang dapat memberikan gambaran tentang bagaimana masyarakat bereaksi terhadap berbagai isu pemilu. Teknik ini membantu mengidentifikasi tren sentimen dominan, yang dapat digunakan untuk mengantisipasi potensi masalah sosial atau politik yang mungkin muncul pasca pemilu.

Analisis sentimen memberikan wawasan kritis tentang bagaimana masyarakat bereaksi terhadap hasil Pemilu 2024, yang dapat mempengaruhi stabilitas politik dan sosial di Indonesia. Dengan memahami sentimen publik, para pembuat kebijakan dan politisi dapat membuat keputusan yang lebih informatif dan strategis untuk mengelola opini publik dan merespons isu-isu kritis. Selain itu, penelitian ini juga berkontribusi pada pengembangan metode analisis sentimen yang lebih akurat dan efisien, yang dapat diterapkan dalam berbagai konteks sosial dan politik lainnya di masa depan. Analisis sentimen membantu dalam memahami dan memperoleh informasi yang relevan dari data tekstual, sehingga dapat membantu dalam mengatasi rumor kontroversial dengan lebih baik. Penelitian ini penting untuk memastikan proses pemilu yang lebih transparan, adil, dan akuntabel di masa mendatang.

2. Metodologi Penelitian

Penelitian ini dilakukan untuk melihat sentiment masyarakat terkait dengan hasil pemilihan umum 2024. Disini kita akan mengambil data *tweet* dari Twitter menggunakan Teknik Crawling data dengan API Twitter. Tweet dari twitter diambil dari tanggal 15 Februari sampai 30 April 2024 berdasarkan keyword (“nama capres dan cawapres”). Data yang digunakan dikelompokkan berdasarkan bulan, dengan tujuan agar kita bisa melihat bagaimana pendapat masyarakat terkait hasil pemilu dari perolehan suara masing- masing capres dan cawapres perbulannya. Setelah data tweet berhasil didapatkan, selanjutnya dilakukan tahap *preprocessing* data, dimulai dari *cleaning data* seperti penghapusan *hashtag*, *mention*, emotikon, spasi yang berlebih, dan lainnya. kemudian *casefolding*, tokenisasi, *stemming*, normalisasi dan *stopword*. Setelah tahap *preprocessing* selesai, kemudian dilakukan pelabelan data menjadi positif, netral dan negatif menggunakan 4 library yaitu: *lexicon Bahasa Indonesia*, *textblob*, *nlk* dan *vader sentiment*. Tujuan kita menggunakan 4 library tersebut agar kita bisa melihat seberapa akurat *library* tersebut untuk melakukan pelabelan. Selanjutnya tahap *data mining*, kita menggunakan metode algoritma Naïve Bayes untuk mengukur keakuratan pelabelan yang kita lakukan sebelumnya. Lalu tahap terakhir yaitu evaluasi atau pengujian menggunakan *confusion matriks*.

Untuk mempermudah pemahaman tentang alur penelitian, gambar 3.1 dibawah ini yang menjelaskan secara visual tentang proses penelitian.



Gambar 1 Desain Sistem

1. Data Selection

Tahap awal dalam penelitian ini adalah seleksi data, di mana kita menggunakan teknik crawling data untuk mengumpulkan informasi yang relevan dari media sosial terkait dengan pertanyaan penelitian, terutama terkait Pemilihan Umum 2024[4]. Untuk lebih spesifiknya kita menggunakan data dengan keyword (“capres dan cawapres) mulai dari tanggal 15 february sampai 30 april yang dikelompokkan berdasarkan bulan. Library yang digunakan untuk *crawling data* dari twitter yaitu “Tweet Harvest” menggunakan API twitter. Proses untuk melakukan *crawling data* dengan menjalankan kode program menggunakan google colab dan API twitter dari pengguna[5].

2. PreProcessing

Preprocessing merupakan tahap persiapan data yang bertujuan agar mempermudah proses pengolahan data. Preprocessing memfokuskan pada data *cleaning & cleansing*, termasuk menghilangkan noise di data, mengatasi struktur data yang tidak baik, dan informasi yang hilang. Sebelum melakukan klasifikasi, data perlu diatur dan dibersihkan selama tahap *pra-processing* melibatkan langkah-langkah berikut ini[6]:

a) Cleaning

Proses pembersihan data melibatkan penghapusan emotikon atau simbol tambahan yang tidak relevan atau tidak diperlukan.

b) *Case Folding*

Dalam langkah pertama dari prosedur ini, dilakukan upaya untuk mengubah teks menjadi huruf kecil.

c) *Tokenizing*

Tahap berikutnya adalah tokenisasi, di mana kalimat akan dipisahkan menjadi kata-kata secara individu. Dengan menggunakan metode ini, kalimat dapat dipecah menjadi unit-unit kata yang terpisah, memungkinkan untuk analisis lebih lanjut pada tingkat kata..

d) *Stemming*

Prosedur yang dilakukan untuk menghilangkan imbuhan adalah dengan menghapus imbuhan setelah melakukan tokenisasi. Dalam langkah ini, imbuhan pada hasil token akan dihapus untuk mendapatkan bentuk dasar atau kata dasar dari setiap kata yang ada.

e) *Normalization*

Normalisasi disini untuk mengembalikan sebuah kata ke bentuk standar. Contohnya: kata tdk menjadi tidak, yg menjadi yang.

f) *Stopword Removal*

kata-kata umum yang sering muncul, yang tidak memberikan informasi penting (yang biasanya tidak diacuhkan atau dibuang misalnya dalam proses pembuatan indeks atau daftar kata).

3. Pelabelan Data

Setelah proses *preprocessing*, kita melakukan proses pelabelan. Pelabelan analisis sentimen dalam metodologi penelitian merujuk pada proses memberikan label atau kategori tertentu kepada teks berdasarkan sentimen atau emosi yang terkandung di dalamnya. Ini merupakan langkah awal penting dalam banyak penelitian tentang analisis sentimen, di mana kita ingin mengidentifikasi dan memahami sentimen yang terkandung dalam teks. Proses pelabelan ini digunakan untuk menentukan apakah kalimat atau teks tersebut merupakan sentiment positif, netral, atau negative[7]. Pelabelan dalam penelitian ini menggunakan 4 library yaitu: lexicon Bahasa Indonesia, textblob, nltk, dan vader sentiment[8][9][10][11]. Selain dari *library* lexicon Bahasa Indonesia, hasil dari *preprocessing* kita terjemahkan terlebih dahulu karena *library* tersebut menggunakan Bahasa Inggris. Kemudian baru kita melakukan pelabelan dengan masing-masing *library*. Dengan tujuan agar kita bisa mengetahui perbedaan pelabelan dari masing-masing *library*.

Pelabelan analisis sentimen merupakan langkah awal yang penting dalam metodologi penelitian analisis sentimen, karena kualitas label sentimen yang diberikan akan mempengaruhi hasil akhir dari penelitian. Oleh karena itu, proses pelabelan harus dilakukan dengan hati-hati dan konsisten agar menghasilkan dataset yang valid dan dapat dipercaya untuk analisis lebih lanjut.

4. *Data Mining*

Langkah berikutnya adalah data mining, dimana data mentah diubah menjadi informasi yang berharga. Pada tahap ini, tujuannya adalah untuk menemukan dan menganalisis pola tersembunyi dalam jumlah besar data guna mendapatkan wawasan yang berguna dan bermanfaat[12]. Salah satu algoritma klasifikasi yang digunakan adalah Naive Bayes berdasarkan Teorema Bayes[13]. Setelah melakukan pelabelan pada data teks menggunakan berbagai metode, kita bisa menggunakan algoritma klasifikasi seperti Naive Bayes untuk mengukur akurasi dari pelabelan tersebut. Langkah ini melibatkan penggunaan data latih yang sudah diberi label dan data uji untuk melatih dan menguji model Naive Bayes. Dengan menggunakan data latih yang telah dilabeli, model Naive Bayes dapat mempelajari pola dari fitur-fitur teks yang dikaitkan dengan label tertentu. Kemudian, model tersebut diuji pada data uji untuk mengukur seberapa baik model tersebut dalam memprediksi label yang benar. Hasil akurasi dari model Naive Bayes dapat memberikan gambaran tentang seberapa baik pelabelan yang dilakukan dengan metode sebelumnya.

5. *Evaluation*

Pada langkah terakhir, model yang telah dibentuk akan dievaluasi untuk memastikan konsistensi dengan hipotesis sebelumnya. Untuk mengukur akurasi algoritma, kita menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk menggambarkan prediksi model dan seberapa baik metode tersebut diterapkan. Dengan menggunakan *confusion matrix*, kita dapat menganalisis kinerja model dalam empat dimensi dan mendapatkan pemahaman yang lebih jelas tentang performa yang dicapai:

- a Akurasi: Untuk mengukur sejauh mana nilai yang diharapkan dan nilai aktual memiliki kemiripan satu sama lain.
- b Presisi : Untuk mengevaluasi keakuratan atau presisi suatu model berdasarkan prediksi positifnya. Presisi adalah metrik yang berguna saat model memiliki tingkat Positif Palsu yang tinggi.
- c Recall : Untuk menilai berapa banyak nilai Positif Aktual yang telah dikenali sebagai True Positif oleh model melalui pelabelan. Recall adalah metrik yang

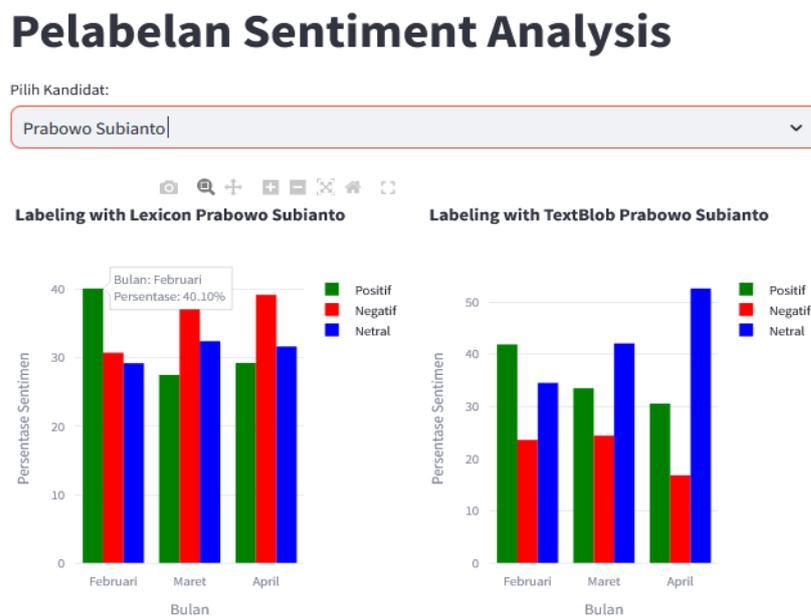
digunakan saat terdapat tingkat False Negatif yang tinggi. Recall menjadi metrik penting dalam menentukan model yang terbaik.

- d F-Measure: Perbandingan antara presisi dan recall dapat diukur menggunakan F1-score, yang merupakan perolehan hasil rata-rata tertimbang dari kedua metrik tersebut. F1-score memberikan gambaran keseluruhan tentang keseimbangan antara presisi dan recall dalam penilaian kinerja model[14].

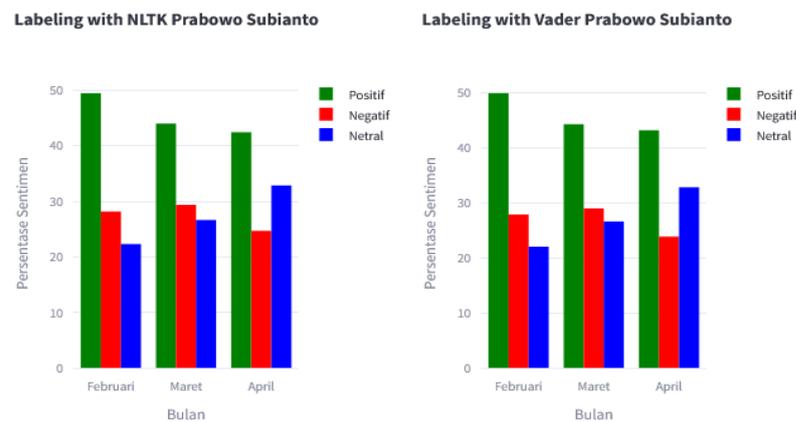
3. Hasil dan Pembahasan

A Perbandingan 4 library untuk pelabelan

Pelabelan menggunakan 4 library tersebut memiliki hasil yang berbeda, bisa dilihat pada Gambar 2 dan 3 berikut ini:



Gambar 2 Dashboard Pelabelan menggunakan Lexicon



Gambar 3 Dashboard Pelabelan menggunakan Vader

Dari grafik untuk kandidat calon presiden Prabowo, sentiment positif paling tinggi pada library vader sentiment bulan februari yaitu sebesar 50%, sentiment netral paling tinggi pada library Textblob bulan April yaitu sebesar 52,61%, dan sentiment negatif paling tinggi pada library Lexicon bulan maret sebesar 40.11%.

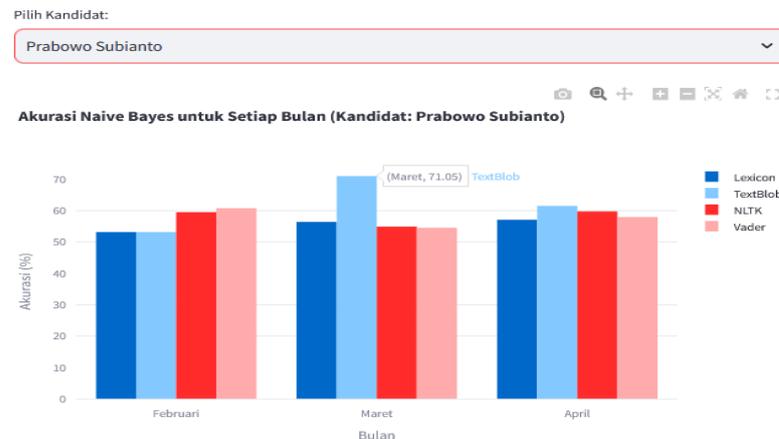
Masing-masing *library* memiliki daftar kata atau kamus data yang sudah diberi bobot atau polarity. Bobot yang diberikan berbeda dari masing-masing *library*. contohnya: polarity kata love pada library textblob adalah 0.5 sedangkan pada library vader sentiment adalah 0.6696. kata educational pada textblob adalah 0.1 sedangkan pada vader sentiment adalah 0.0. Jadi masing-masing *library* memiliki cara sendiri untuk melakukan pembobotan kata. Disini kita hanya menggunakan *library* yang sudah disediakan.

Bobot kata lebih dari 0 itu dikategorikan positif, bobot kata kurang dari 0 dikategorikan negatif, dan bobot kata sama dengan 0 dikategorikan netral. Cara kerja masing-masing *library* untuk melakukan pelabelan yaitu dengan menjumlahkan seluruh kata yang sudah diberi bobot dalam sebuah kalimat lalu di rata-ratakan. Maka hasil dari rata-rata tersebutlah yang digunakan untuk klasifikasi atau pelabelan untuk analisis sentiment.

Tujuan kita untuk membuat pelabelan menggunakan 4 *library* sekaligus yaitu untuk mengetahui dan membandingkan *library* yang paling akurat untuk melakukan pelabelan pada analisis sentiment untuk data tweet twitter.

B Hasil Akurasi dari Algoritma Naïve Bayes

Accuracy Sentiment Analysis



Gambar 4 Hasil Accuracy

Dari grafik accuracy Gambar 4 menggunakan algoritma naïve bayes ditarik kesimpulan bahwa akurasi tertinggi menggunakan *library* Textblob sebesar 71.05%.

Setelah melakukan pelabelan menggunakan *library* lexicon Bahasa Indonesia, TextBlob, NLTK, dan Vader sentiment selanjutnya kita mengukur akurasi atau keakuratan dari pelabelan yang dilakukan. Hasil akurasi untuk 4 pelabelan tersebut berbeda-beda. Ada beberapa hal yang mempengaruhi akurasi yaitu:

- 1) Pertama jumlah data, semakin banyak data yang diolah, maka akurasi akan semakin tinggi, dan semakin kurangnya data maka akurasi akan semakin rendah.
- 2) Kedua pembagian *data training* dan *data testing*. Jika kita membagi data 80% untuk training dan 20% untuk testing dan membagi data tersebut secara acak saja, maka itu membuat hasil akurasi menurun. Sedangkan ketika kita membagi data tersebut masing-masing label itu sama persisnya, maka itu akan sedikit meningkatkan akurasi dari pengujian data (ini bersifat opsional tergantung dari datanya, terkadang menerapkan *stratify* pada data yang lain, akurasi makin menurun). Jadi *stratify* itu digunakan tergantung dengan kebutuhan dan datanya (apakah data seimbang atau tidak). Jika seimbang tidak butuh menggunakan *stratify*, tapi jika data tidak seimbang maka dibutuhkan *stratify*.
- 3) Ketiga *library* yang digunakan, 4 *library* yang kita gunakan memiliki cara tersendiri untuk melakukan pelabelan. Maka hal itu juga yang mempengaruhi akurasi dari pengujian data. Untuk contoh kita akan menghitung akurasi dari pelabelan Cawapres Muhaimin bulan April. Lexicon: 0.60%, TextBlob: 0.72%, NLTK: 0.64%, VaderSentiment: 0,62%. *Library* yang memiliki akurasi yang paling tinggi yaitu TextBlob yaitu: 0.72 %. Hal itu tidak selalu sama dalam semua kandidat, textblob tidak selalu memiliki akurasi yang tinggi dalam

pengujian, itu tergantung dari model data yang akan digunakan, jumlah data, pembagian data dan *preprocessing data* yang dilakukan.

C Hasil Confusion Matrix

Hasil *Precision*, *recall*, dan *F1-score* pada *confusion matrix* hasilnya hampir sama. Ada beberapa hal yang menyebabkan hal tersebut terjadi yaitu:

1) Keseimbangan Kelas :

Jika dataset seimbang, artinya jumlah instance untuk setiap kelas hampir sama, maka *precision* dan *recall* untuk setiap kelas juga cenderung seimbang. Dalam kasus ini, *F1-score*, yang merupakan rata-rata harmonik dari keduanya, akan mencerminkan nilai yang mirip.

2) Kinerja Model yang Konsisten :

Jika model memiliki kinerja yang baik dan konsisten dalam mengklasifikasikan positif dan negatif, maka jumlah FP dan FN akan rendah. Dengan FP dan FN yang rendah, nilai *precision* dan *recall* akan tinggi dan hampir sama, sehingga *F1-score* juga akan mirip.

3) Distribusi Data yang Homogen :

Jika data tes memiliki distribusi yang sangat mirip dengan data latih dan model telah dilatih dengan baik pada data tersebut, maka model akan memiliki kinerja yang baik pada data tes. Hal ini menyebabkan *precision*, *recall*, dan *F1-score* menjadi tinggi dan mirip.

4) Perlakuan yang sama pada kesalahan

Jika model melakukan kesalahan yang konsisten di semua kelas (misalnya, jika model cenderung *overfit* atau *underfit* secara merata), maka *precision* dan *recall* akan mencerminkan ini secara proporsional, yang mengarah ke *F1-score* yang serupa.

Hasil yang hampir mirip adalah indikator bahwa model memiliki kinerja yang seimbang dalam mendeteksi *true positive* dan menghindari *false positive* serta *false negative*. Ini adalah tanda positif bahwa model bekerja dengan baik pada dataset yang diberikan.

4. Kesimpulan

1. Hasil pelabelan mendapatkan hasil yang bervariasi dari masing-masing *library*. Yang mempengaruhi hasil pelabelan yaitu karakteristik dataset, konteks kalimat (secara keseluruhan, konteks kalimat yang mengakibatkan kesalahan dalam penilaian sentiment), kelengkapan kamus kata (semakin lengkap dan akurat kamus kata yang digunakan, semakin baik hasil pelabelan), dan kualitas dan kuantitas data pelatihan. Semakin baik dan representatif data pelatihan, semakin baik hasil pelabelan.

2. Hasil sentiment dari 6 kandidat pemilu ini dapat disimpulkan bahwa persentase sentiment positif paling banyak yaitu kandidat Ganjar Pranowo yaitu sebesar 63.78%, kedua Mahfud MD yaitu 58.91%, ketiga Anies yaitu 48.92%, keempat Muhaimin 46.35%, kelima Prabowo yaitu 45.84%, keenam Gibran yaitu 43.74%. Sentiment negative paling banyak yaitu Anies Baswedan yaitu 28.68%, Muhaimin Iskandar yaitu 28.65%, Prabowo Subianto yaitu 26.95%, Gibran Rakabuming Raka yaitu 26.66%, Ganjar Pranowo yaitu 23.73%, dan Mahfud yaitu 22.40%. Jadi disimpulkan bahwa sentiment masyarakat terkait hasil pemilu 2024 banyak sentiment positif kepada kandidat Ganjar Pranowo dan negative kepada Anies Baswedan.
3. Hasil akurasi paling tinggi menggunakan algoritma naïve Bayes yaitu menggunakan library lexicon Bahasa Indonesia sebesar 81.54%, kedua menggunakan library vader sentiment sebesar 80.91%, ketiga menggunakan library NLTK sebesar 80.08%, dan keempat menggunakan library TextBlob sebesar 73.54%.

References

- [1] H. Simorangkir dan K. M. Lhaksana, "Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier," *e-proceeding of Englineering*, vol. 5, no. 3, hal. 8131–8140, 2018, [Daring]. Tersedia pada: https://openlibrary.telkomuniversity.ac.id/pustaka/files/144621/jurnal_eproc/analisis-sentimen-pada-twitter-untuk-games-online-mobile-legends-dan-arena-of-valor-dengan-metode-na-ve-bayes-classifier.pdf
- [2] E. A. Rostiati, "Manajemen Risiko Pengadaan Dan Distribusi Logistik Pemilu Serentak 2024 Berbasis Coso Enterprise Risk Management," *Elect. Gov. J. Tata Kelola Pemilu Indones.*, vol. 5, hal. 89, 2023, [Daring]. Tersedia pada: <http://www.nber.org/papers/w16019>
- [3] M. Nurkamiden, "SiRekap : Tantangan dan Potensi Kekeliruan Proses Rekapitulasi Pemilu Serentak di Indonesia," *Sosiol. J. Penelit. dan Pengabd. Kpd. Masy.*, vol. 1, no. 2, hal. 101–110, 2024, [Daring]. Tersedia pada: <http://ejurnal.fis.ung.ac.id/index.php/sjppm/about%0ASiRekap>
- [4] R. Vindua dan A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, hal. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [5] R. Gelar Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, hal. 55–60, 2023, doi: 10.47233/jteksis.v5i1.750.
- [6] M. Alfonso dan D. B. Rarasati, "Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm," *JISA(Jurnal Inform. dan Sains)*, vol. 6, no. 2, hal. 110–115, 2023, doi: 10.31326/jisa.v6i2.1714.
- [7] T. D. Putra, E. Utami, dan M. P. Kurniawan, "Analisis Sentimen Pemilu 2024 dengan Naïve Bayes Berbasis Particle Swarm Optimization (PSO)," *Explore*, vol. 13, no. 1, hal. 1–5, 2023, doi: 10.35200/ex.v11i2.13.
- [8] S. Anggina, N. Y. Setiawan, dan F. A. Bachtiar, "Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF Pada Formaggio Coffee and Resto," *is Best Account. Inf. Syst. Inf. Technol. Bus. Enterp. this is link OJS us*, vol. 7, no. 1, hal. 76–90, 2022, doi: 10.34010/aisthebest.v7i1.7072.
- [9] M. U. Albab, Y. Karuniawati P, dan M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," *J. Transform.*, vol. 20, no. 2, hal. 1–10, 2023, [Daring]. Tersedia pada: <https://journals.usm.ac.id/index.php/transformatika/page1>

- [10] R. Chandra, K. Gunadi, dan S. Ananda, “Aplikasi Sentiment Analysis terhadap Trend Cryptocurrency pada Platform Twitter Menggunakan Library Textblob sebagai Alat Bantu Berinvestasi,” *J. Infra*, vol. 10, no. 2, hal. 397–403, 2022, [Daring]. Tersedia pada: <https://publication.petra.ac.id/index.php/teknik-informatika/article/viewFile/12809/11109>
- [11] S. Ernawati dan R. Wati, “Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon,” *J. Buana Inform.*, vol. 15, no. 1, hal. 40–49, 2024.
- [12] S. Puad, G. Garno, dan A. Susilo Yuda Irawan, “Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, hal. 1560–1566, 2023, doi: 10.36040/jati.v7i3.6920.
- [13] A. Perdana, A. Hermawan, dan D. Avianto, “Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 2, hal. 195–200, 2022, doi: 10.32736/sisfokom.v11i2.1412.
- [14] P. Romadloni, B. Adhi Kusuma, dan W. Maulana Baihaqi, “Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, hal. 622–628, 2022, doi: 10.36040/jati.v6i2.5238.