

KLASIFIKASI BERITA HOAX TERKAIT PEMILIHAN UMUM PRESIDEN REPUBLIK INDONESIA TAHUN 2024 MENGUNAKAN NAÏVE BAYES DAN SVM

CLASSIFICATION OF HOAX NEWS RELATED TO THE GENERAL ELECTION OF THE PRESIDENT OF THE REPUBLIC OF INDONESIA IN 2024 USING NAÏVE BAYES AND SVM

Bahtiar Imran^{*1}, Muh Nasirudin Karim², Nur Isna Ningsih³

*Email: bahtiarimranlombok@gmail.com

¹Rekayasa Perangkat Lunak, Universitas Teknologi Mataram, Mataram, Indonesia

^{2,3}Teknik Informatika, Universitas Teknologi Mataram, Mataram, Indonesia

Abstrak

Pemilihan Umum Presiden di Indonesia merupakan salah satu peristiwa penting dalam kehidupan berpolitik negara. Peristiwa ini selalu mendapat perhatian publik yang besar dari segenap masyarakat. Seperti pada era digital ini, akses berita sangat mudah diakses oleh setiap orang, dan hal ini pula dimanfaatkan oleh beberapa oknum tak bertanggung jawab guna mendapat keuntungan bagi diri sendiri, kelompok, maupun golongan tertentu dengan cara menyebarkan berita *hoax* mengenai isu pilpres di Indonesia. Karenanya, penelitian ini bertujuan mengimplementasikan metode klasifikasi berita hoaks terkait Pilpres 2024 menggunakan algoritma *Naive Bayes* serta *Support Vector Machine* (SVM). Evaluasi dilakukan dengan membandingkan kinerja kedua algoritma dalam hal akurasi, presisi, *recall*, serta *F1-score*. Hasil menunjukkan bahwa *Naive Bayes* mencapai akurasi sebesar 97% dengan presisi 94%, *recall* 100%, dan *F1-score* 97%. Sementara itu, SVM memiliki akurasi 95%, presisi 94%, *recall* 97%, serta *F1-score* 95%. Kesimpulan didapatkan bahwa kedua algoritma memiliki kinerja yang layak dalam klasifikasi berita *hoax non hoax* pemilihan umum presiden di Indonesia tahun 2024.

Kata kunci: *hoax, naïve bayes, pilpres, support vector machine.*

Abstract

Presidential elections in Indonesia are one of the most important events in the country's political life. This event always receives great public attention from the whole community. As in this digital era, access to news is very easy to access by everyone, and this is also utilized by some irresponsible people to benefit themselves, groups, and certain groups by spreading hoax news about presidential election issues in Indonesia. Therefore, this research aims to implement a hoax news classification method related to the 2024 presidential election using the Naive Bayes algorithm and Support Vector Machine (SVM). The evaluation is done by comparing the performance of both algorithms in terms of accuracy, precision, recall, and F1-score. The results showed that Naive Bayes achieved 97% accuracy with 94% precision, 100% recall, and 97% F1-score. Meanwhile, SVM has 95% accuracy, 94% precision, 97% recall, and 95% F1-score. It can be concluded that both algorithms have decent performance in the classification of non hoax news of the presidential general election in Indonesia in 2024.

Keywords: *hoax, naïve bayes, pilpres, support vector machine*

I. PENDAHULUAN

Indonesia adalah negara demokratis dengan jumlah penduduk terbesar keempat di dunia

(274.790.244), dan menganut sistem demokrasi. Di Indonesia, pemilihan umum diadakan setiap lima tahun sekali. Pemilihan presiden kedelapan akan diadakan pada tahun 2024 [1]. Menyalurkan opini

publik melalui pemilihan umum yang teratur adalah salah satu prinsip utama dari setiap pemerintahan yang demokratis. Indonesia merupakan salah satu penerapan demokrasi adalah melalui pemilihan umum. Pemilihan umum yang diselenggarakan untuk memilih wakil rakyat dan pejabat publik lainnya adalah salah satu cara untuk menerapkan demokrasi [2]. Salah satu tonggak penting dalam membangun demokrasi di Negara Kesatuan Republik Indonesia adalah pemilihan presiden yang akan diadakan tahun 2024. Media sosial adalah alat yang dapat digunakan oleh para calon presiden untuk menyebarkan pesan kampanye politik mereka [3]. Seiring dengan berakhirnya masa jabatan Presiden dan Wakil Presiden Indonesia saat ini, tahun 2024 akan dikenang oleh seluruh rakyat Indonesia sebagai tahun pesta demokrasi terbesar, dalam menentukan Wakil Presiden dan Presiden Indonesia [4]. Terkait dengan pesta demokrasi terbesar ini banyak sekali berita hoax yang tersebar terkait dengan pemilihan umum maupun terhadap pasangan calon di media sosial. Platform media sosial memungkinkan pengguna untuk berbagi, berjejaring, berkomunikasi, dan melakukan banyak hal lainnya. Melalui penggunaan teknologi berbasis situs web atau aplikasi, media sosial dapat mengubah percakapan menjadi dialog interaktif, menghidupkan percakapan dan membuatnya menjadi interaktif. *YouTube, Facebook, blog, Twitter*, dan *platform* lainnya adalah beberapa contoh media sosial. Media sosial terbuka bagi semua pengguna untuk mengunggah hal-hal yang positif maupun negatif. Mereka juga dapat dengan bebas menyebarkan informasi palsu atau informasi yang tidak sesuai dengan kebenaran, seperti berita palsu [5]. *Hoax* didefinisikan sebagai kebohongan atau informasi yang palsu, fitnah, dan tidak memiliki pola yang jelas. Berita hoaks biasanya dibuat dengan sengaja dengan tujuan untuk menyebarkan ujaran kebencian atau propaganda yang memfitnah individu atau organisasi tertentu [6].

Penelitian yang terkait dengan berita hoaks pada pemilihan presiden 2024 masih belum terlalu banyak dan penelitian terdahulu menggunakan metode yang berbeda dalam mengklasifikasi berita hoaks terkait dengan pemilihan presiden 2024, seperti pada penelitian [3] menggunakan *naive bayes* berbasis *Particle Swarm Optimization* (PSO) dan mendapatkan hasil dengan menggunakan *Naive Bayes* tanpa PSO nilai akurasi sebesar 73,67%. Sedangkan untuk algoritma *Naive Bayes* berbasis PSO nilai akurasi sebesar 78,33%. Penelitian [1] menggunakan *naive bayes* untuk melakukan sentimen analisis terkait dengan bakal calon presiden 2024, dan

mendapatkan hasil akurasi menggunakan algoritma *naive bayes classifier* sebesar 78% yang di dapat dengan menggunakan perbandingan 10% data uji serta 90% data latih. Penelitian [2] menggunakan python untuk melakukan analisis sentimen pada pemilu 2024 dengan mengambil data dari *twitter*, hasil yang didapatkan persentase sentimen kelas positif sebesar 40%, netral 52% dan negatif 8% tentang Pemilu tahun 2024. Pada [4] melakukan sentimen analisis pada calon presiden yang potensial dengan memanfaatkan data dari *twitter*, hasil yang didapatkan Evaluasi algoritma klasifikasi menunjukkan *Support Vector Machine* (SVM) dengan akurasi 84%, *Bernoulli Naive Bayes* dengan 77%, dan Regresi Logistik dengan 84%.

Penelitian yang telah dilakukan sebelumnya dengan topik yang sama, belum ada penelitian yang menggunakan *naive bayes* dan *svm* untuk melakukan klasifikasi terhadap berita hoaks terkait dengan pemilihan umum 2024. Berdasarkan permasalahan tersebut perlu sebuah metode untuk mendeteksi sebuah berita hoax atau fakta. Penelitian ini menggunakan Metode *Naive Bayes* Metode SVM, Dengan penggunaan metode ini diharap dapat memberikan model yang tepat dengan tingkat akurasi yang tinggi, sehingga dapat mencegah beredarnya informasi maupun berita hoax ditengah masyarakat menjelang pemilihan umum presiden tahun 2024 di Indonesia. Alasan pemilihan metode *naive bayes* dikarenakan metode ini mampu memberikan hasil akurasi klasifikasi yang tinggi [7], sedangkan SVM merupakan metode klasifikasi dengan hasil yang baik [8].

II. TINJAUAN PUSTAKA

A. Berita Hoaks

Selain menjadi alat untuk mempengaruhi opini publik, berita palsu juga melayani kepentingan individu atau kelompok tertentu. kepentingan individu atau kelompok tertentu. Selain itu, mereka yang menyebarkan berita palsu yang kontroversial dapat memperoleh keuntungan dengan menarik lebih banyak pengunjung ke situs *web* yang diiklankan. Pada kenyataannya, korban berita palsu dapat disesatkan dan dirugikan sebagai akibatnya. Dampak yang merugikan termasuk kerugian material dan moral seperti reputasi yang rusak, serta potensi permusuhan pembaca untuk tumbuh hingga ke titik di mana korban berada dalam bahaya [9].

B. Naive Bayes

Naive Bayes menggunakan teorema *Bayes* untuk membuat klasifikasi probabilitas langsung. Sederhana, dapat diandalkan, dan otonom,

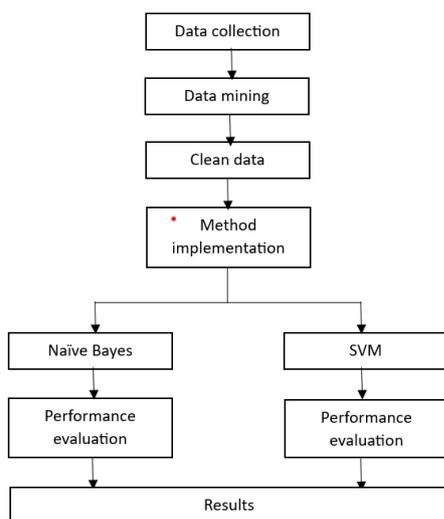
pendekatan ini bersifat mandiri. $P(d|c)$ adalah apa yang dituliskan oleh *Naïve Bayes* jika "d" adalah vektor fitur input dan "c" adalah label kelas. Notasi ini menunjukkan kemungkinan bahwa kelas "c" akan ditemukan setelah fitur "d" ditemukan. $P(c)$ dikenal sebagai probabilitas sebelumnya, dan $P(d|c)$ dikenal sebagai probabilitas kemungkinan pertama. Probabilitas akhir dipelajari selama proses pelatihan menggunakan data dari set pelatihan [10].

C. Support Vector Machine

Dorongan Model pembelajaran yang diawasi yang dikenal sebagai mesin vektor /SVM dan algoritme pembelajaran yang menyertainya memeriksa data yang digunakan dalam studi regresi dan klasifikasi. dengan menawarkan set data pelatihan yang dibagi menjadi dua kategori. SVM mampu melatih dan membangun model untuk prediksi dan klasifikasi [11].

III. METODE

Metode penelitian yang digunakan yakni Metode *Naive Bayes* serta perbandingan berupa Metode SVM serta input data ialah teks. Setelah itu, teks tersebut akan dilakukan proses *data mining* seperti *tokenization*, *stemming*, *stopword removal*, *case folding* dan pembobotan kata memakai TF-IDF [12]. Tahap selanjutnya ialah memasukkan metode klasifikasi berita *hoax* memakai *Naive Bayes* serta perbandingnya memakai Metode SVM. Kemudian, akan dilakukan evaluasi kinerja menggunakan metode akurasi, presisi, recall, serta *F-1 Score*. Hasil akhir dari *output* ini berupa perbandingan akurasi dari Metode *Naive Bayes* serta SVM. Untuk tahapan-tahapan yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Tahap Perancangan Penelitian

A. Text Mining

Text mining yakni proses guna mendapatkan informasi yang memiliki kualitas tinggi dari teks. Informasi yang memiliki kualitas tinggi umumnya diperoleh dengan mengawasi tren serta pola, lalu mempelajarinya. Dalam proses *text mining* ada pembobotan kata yang berguna demi memberi bobot atau nilai atas term yang ada dalam sebuah dokumen. Nilai yang diberikan untuk *term* menyesuaikan dengan metode yang dipakai [6]. Pada *text mining* ini dilakukan tahapan *Text Preprocessing*. *Text Preprocessing* yakni fase pertama dari *Data Mining*. Pada fase tersebut dijalankan proses pengolahan data agar data siap dalam proses klasifikasi [12] Tahap *Text Mining* tersebut dapat menghasilkan kalimat berita *non hoax* serta *hoax* diseleksi serta tersimpan pada dokumen dengan format *xlsx*. Tahapan-tahapan dalam *Text Mining* meliputi *tokenisasi*, *Case folding*, *Stopword Removal*, *Stemming*, serta pembobotan kata menggunakan TF-IDF.

1) Tokenisasi

Tokenisasi yakni cara memecahkan teks jadi bagian-bagian kecil (*bigrams*, kalimat, kata). Mencakup: *Punctuation* serta *Word* bisa mengelompokkan teks dari kata serta mengecualikan simbol tanda baca (tanpa menghapusnya) [13].

2) Case Folding

Case Folding yakni mengubah seluruh huruf jadi huruf kecil. Semisal: mengubah "CASE" jadi "case", "FOLDING" jadi "folding", serta seterusnya [14]. *Case Folding* merubah semua huruf pada suatu dokumen jadi huruf kecil yakni terbatas pada huruf a-z. Adapun yang tidak termasuk huruf a-z akan dianggap delimiter serta dihilangkan [15].

3) Stopword Removal

Stopword removal yakni proses menghilangkan kata-kata penghubung umum (*stopword*) dari sebuah teks dalam analisis bahasa alami atau pemrosesan teks. Tujuan dari menghapus kata yakni untuk menghapus kata-kata yang dianggap tak penting, proses tersebut bisa memangkas dimensi ruang yang terlihat berat. Kata-kata umum pada dokumen teks semisal pakaian, kata depan dan kata benda dan lain-lain, yang tidak memberi arti pada dokumen tersebut [16].

4) Stemming

Stemming yakni proses memetakan serta menguraikan bentuk kata jadi bentuk kata dasar. Singkatnya yakni proses merubah kata yang memiliki imbuhan jadi kata dasar. Tahapan tersebut memiliki peranan penting pada tahap *Text Mining*. Hasil *Stemming* yang baik bisa berpengaruh terhadap kualitas pengaplikasian *Text Mining* [17].

5) TF – IDF

Term Frequency Inverse Document Frequency (TF-IDF) yakni metode pembobotan yang mengkombinasikan metode IDF serta TF. Metode ini diusulkan oleh Salton sebagai suatu gabungan metode yang bisa memberi performansi yang lebih baik, terutama untuk memperbaiki nilai *precision* serta *recall* [18] *Term Frequency Inverse Document Frequency* atau TF-IDF yakni metode guna mengevaluasi bobot seluruh kata dalam dokumen serta beberapa dokumen, bobot tersebut jadi presentasi penting dalam suatu kata pada dokumen. Semakin besar bobotnya, maka peranan kata tersebut menjadi lebih penting dalam presentasi suatu dokumen [12].

B. Metode Naïve Bayes

Naïve Bayes yakni salah satu cara *machine learning* yang memakai perhitungan probabilitas. Konsep dasar yang dipakai oleh *Naive Bayes* yakni *Teorema Bayes*, yakni menjalankan klasifikasi dengan perhitungan nilai probabilitas hasil *Naive Bayes Classifier* ataupun NBC yakni proses mengklasifikasikan probabilitas sederhana yang mengacu pada *Bayes Theory*. Teori tersebut mengungkapkan bahwasanya kemungkinan adanya sebuah peristiwa sama dengan probabilitas intinsik (dihitung dari data yang tersedia saat ini) dikalikan probabilitas bahwasanya peristiwa semacam itu akan ada lagi di masa mendatang (berdasarkan pengetahuan yang adanya di masa lalu) [19].

Naïve Bayes merupakan salah satu algoritma klasifikasi yang sederhana tapi mempunyai tingkat keakurasian yang tinggi, akan tetapi *Naive Bayes* mempunyai kekurangan yakni sangat sensitive pada pemilihan fitur hingga bisa mempengaruhi hasil keakurasiannya. Metode klasifikasi berbasis fitur yang dikembangkan pada studi-studi tersebut memberikan tingkat keakurasian yang baik [20]. Secara umum, *Teorema Bayes* digambarkan dalam persamaan berikut:

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

Keterangan :

- x : Data dengan *class* yang belum diketahui
- H : Hipotesis data x merupakan suatu *class* spesifik
- $P(H|x)$: Probabilitas hipotesis H berdasarkan kondisi x (*posterior probability*).
- $P(x|H)$: Probabilitas x berdasarkan hipotesis H

$P(x)$: Probabilitas dari x

C. Metode Support Vector Machine

Support Vector Machine (SVM) yakni suatu model pembelajaran yang diawasi, memakai algoritma terkait untuk menganalisis data dalam konteks analisis klasifikasi dan regresi. SVM melakukan pelatihan dengan menyediakan *dataset* dan menentukan dua kategori berbeda di dalamnya. Proses pelatihan SVM melibatkan pembentukan sebuah model, yang kemudian digunakan untuk melakukan prediksi dan klasifikasi [21]. SVM mencapai pelatihan dengan membuat garis, yang dikenal sebagai *hyperplane*, serta membentuk *margin* untuk setiap label dengan menentukan *vektor* terdekat dengan *margin* untuk kedua label. *Vektor* tersebut selanjutnya disebut sebagai *support vector* [11]. Adapun bentuk umum dari SVM digambarkan dalam persamaan berikut:

$$y(x) = wT\phi(x) + b \quad (2)$$

Keterangan:

- x : Vektor input
- w : Parameter bobot
- $\phi(x)$: Fungsi basis
- b : Suatu bias

D. Evaluasi Kinerja

Evaluasi kinerja dilakukan penghitungan hasil nilai performa yang menggunakan metode akurasi, presisi, *recall* serta *F-1 score* dalam dataset pemilihan umum presiden Indonesia tahun 2024.

Performa yakni bentuk pekerjaan, perbuatan, tindakan yang sudah tercapai ataupun terlaksana, Kinerja dari klasifikasi bisa dinilai dari penghitungan performa nilai akurasi, presisi, *recall*, *F-1 Score* [22].

1) Akurasi

Metrik ini mengevaluasi seberapa jauh model bisa secara benar mengklasifikasikan sampel secara menyeluruh. Akurasi dinilai dengan membagi prediksi yang benar dengan jumlah total sampel. Rumus persamaan akurasi digambarkan pada persamaan (1).

2) Presisi

Presisi mengevaluasi seberapa jauh model memberi prediksi yang benar secara positif. Ini merupakan rasio antara jumlah prediksi benar positif (*True Positive*) serta prediksi positif (*True Positive + False Positive*). Rumus persamaan presisi digambarkan pada persamaan (2).

3) Recall

Recall mengevaluasi seberapa jauh model mampu menemukan ataupun mengenali semua sampel positif yang ada. Ini merupakan rasio antara jumlah prediksi benar positif (*True Positive*) serta jumlah total sampel positif (*True Positive + False Negative*). Rumus persamaan *recall* digambarkan pada persamaan (3).

4) F-1 Score

F1-Score adalah ukuran kombinasi dari presisi dan *recall*. Ini memberikan rata-rata harmonik dari kedua metrik tersebut dan berguna saat kita ingin mencapai keseimbangan antara presisi dan *recall*. Rumus persamaan F-1 Score digambarkan pada persamaan (4). Persamaan Akurasi, Presisi, *Recall*, serta F-1 Score ditulis dalam persamaan berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Precision (p) = \frac{TP}{TP+FP} \quad (4)$$

$$Recall (R) = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = \frac{2*Precision*recall}{precision+recall} \quad (6)$$

Keterangan :

- TP : Jumlah kasus positif yang diprediksi dengan benar oleh model.
- TN : Jumlah kasus negatif yang diprediksi dengan benar oleh model.
- FP : Jumlah kasus negatif yang salah diprediksi sebagai positif oleh model.
- TN : Jumlah kasus positif yang salah diprediksi sebagai negatif oleh model.

IV. HASIL DAN PEMBAHASAN

Penelitian ini membahas tentang bagaimana proses memperoleh dataset terkait dengan berita *hoax* dan *non hoax* pada pilpres di Indonesia tahun 2024. Adapun dalam memperoleh data *hoax* dan *non hoax* memiliki tahapan-tahapan yang harus dilakukan sehingga data tersebut dapat digunakan sebagai *dataset* dalam penelitian. Adapun beberapa metode yang digunakan dalam mendeteksi atau menemukan hasil berita *hoax* dan *non hoax* pada penelitian ini

menggunakan Metode *Naïve Bayes* dan Metode SVM.

A. Data Mining

1) Tokenization

Proses *tokenization*, semua teks berita dalam dataset dipecah menjadi kata-kata dengan menggunakan spasi sehingga mendapatkan kalimat yang dapat dipahami. Tabel 1 merupakan contoh kalimat sebelum dilakukan proses *Tokenization* dan sesudah dilakukan *Tokenization*.

2) Case Folding

Case Folding merupakan proses dimana semua kata diubah menjadi huruf kecil serta dilakukan penghilangan tanda baca dan angka. Tabel 2 merupakan contoh kalimat sebelum dan sesudah dilakukan *Case Folding*.

3) Stopword Removal

Stopword Removal merupakan proses penghilangan kata-kata penghubung dalam kalimat, serta kata - kata yang dianggap tidak penting. Untuk melihat hasil dari *Stopword Removal* dapat dilihat pada tabel 3.

Tabel 1. Sebelum dan sesudah *Tokenization*

Teks Sebelum Dilakukan Proses Tokenization	Teks Setelah Dilakukan Proses Tokenization
Hasil jajak pendapat yang diselenggarakan Litbang Kompas pada 25 Januari 2023 menunjukkan bahwa Partai Nasdem mendapatkan efek ekor jas dari pencalonan Anies Baswedan sebagai bakal calon prsiden. Pencapaian Nasdem itu lantas mengakibatkan turunnya elektabilitas parpol lain yang menjadi basis pemilih Anies, termasuk Partai Demokrat dan Partai Keadlian Sejahtera (PKS). Sebagai informasi, Nasdem, Demokrat dan PKS adalah partai yang menjatuhkan dukungan kepada Anies sebagai calon presiden pada Pilpres 2024 mendatang. Mereka tengah mejajaki kerja sama politik yang diberi nama Koalisi Perubahan.	Hasil jajak pendapat yang diselenggarakan Litbang Kompas pada 25 Januari 2023 menunjukkan bahwa Partai Nasdem mendapatkan efek ekor jas dari pencalonan Anies Baswedan sebagai bakal calon prsiden. Pencapaian Nasdem itu lantas mengakibatkan turunnya elektabilitas parpol lain yang menjadi basis pemilih Anies, termasuk Partai Demokrat dan Partai Keadlian Sejahtera (PKS). Sebagai informasi, Nasdem, Demokrat dan PKS adalah partai yang menjatuhkan dukungan kepada Anies sebagai calon presiden pada Pilpres 2024 mendatang. Mereka tengah mejajaki kerja sama politik yang diberi nama Koalisi Perubahan.

Tabel 2. Sebelum dan sesudah case folding

Contoh Teks Sebelum Case Folding	Contoh Teks Setelah Case Folding
<p>Hasil jajak pendapat yang diselenggarakan Litbang Kompas pada 25 Januari 2023 menunjukkan bahwa Partai Nasdem mendapatkan efek ekor jas dari pencalonan Anies Baswedan sebagai bakal calon prsiden. Pencapaian Nasdem itu lantas mengakibatkan turunnya elektabilitas parpol lain yang menjadi basis pemilih Anies, termasuk Partai Demokrat dan Partai Keadlian Sejahtera (PKS). Sebagai informasi Nasdem, Demokrat dan PKS adalah partai yang menjatuhkan dukungan kepada Anies sebagai calon presiden pada Pilpres 2024 mendatang. Mereka tengah mejajaki kerja sama politik yang diberi nama Koalisi Perubahan.</p>	<p>hasil jajak pendapat yang diselenggarakan litbang kompas pada januari menunjukkan bahwa partai nasdem mendapatkan efek ekor jas dari pencalonan anies baswedan sebagai bakal calon prsiden pencapaian nasdem itu lantas mengakibatkan turunnya elektabilitas parpol lain yang menjadi basis pemilih anies termasuk partai demokrat dan partai keadlian sejahtera pks sebagai informasi nasdem demokrat dan pks adalah partai yang menjatuhkan dukungan kepada anies sebagai calon presiden pada pilpres mendatang mereka tengah mejajaki kerja sama politik yang diberi nama koalisi perubahan</p>

mereka tengah mejajaki kerja sama politik yang diberi nama koalisi perubahan	
------------------------------------------------------------------------------	--

4) **Stemming**

Stemming merupakan proses merubah kalimat menjadi kata dasar, untuk melihat hasil stemming dapat dilihat pada Tabel 4.

Tabel 4. Sebelum dan sesudah di-stemming

Sebelum di Stemming	Setelah di Stemming
<p>hasil jajak pendapat diselenggarakan litbang kompas januari menunjukkan partai nasdem mendapatkan efek ekor jas pencalonan anies baswedan bakal calon prsiden pencapaian nasdem mengakibatkan turunnya elektabilitas parpol basis pemilih anies partai demokrat partai keadlian sejahtera pks informasi nasdem demokrat pks partai menjatuhkan dukungan anies calon presiden pilpres mejajaki kerja sama politik koalisi perubahan</p>	<p>hasil jajak pendapat selenggara litbang kompas januari tunjuk partai nasdem dapat efek ekor jas calon anies baswedan bakal calon prsiden capai nasdem akibat turun elektabilitas parpol basis pemilih anies partai demokrat partai adli sejahtera pks informasi nasdem demokrat pks partai jatuh dukung anies calon presiden pilpres jajak kerja sama politik koalisi rubah</p>

Tabel 3. Sebelum dan sesudah Stopword Removal.

Sebelum Stopword Removal	Setelah Stopword Removal
<p>hasil jajak pendapat yang diselenggarakan litbang kompas pada januari menunjukkan bahwa partai nasdem mendapatkan efek ekor jas dari pencalonan anies baswedan sebagai bakal calon prsiden pencapaian nasdem itu lantas mengakibatkan turunnya elektabilitas parpol lain yang menjadi basis pemilih anies termasuk partai demokrat dan partai keadlian sejahtera pks sebagai informasi nasdem demokrat dan pks adalah partai yang menjatuhkan dukungan kepada anies sebagai calon presiden pada pilpres mendatang</p>	<p>hasil jajak pendapat diselenggarakan litbang kompas januari menunjukkan partai nasdem mendapatkan efek ekor jas pencalonan anies baswedan bakal calon prsiden pencapaian nasdem mengakibatkan turunnya elektabilitas parpol basis pemilih anies partai demokrat partai keadlian sejahtera pks informasi nasdem demokrat pks partai menjatuhkan dukungan anies calon presiden pilpres mejajaki kerja sama politik koalisi perubahan</p>

5) **TF-IDF**

TF-IDF merupakan proses pemberian bobot kata terhadap kata yang terdapat dalam teks. Pembobotan kata akan bergantung dengan seberapa banyak frekuensi kemunculan kata tersebut di dalam data yang digunakan [12]. Contoh hasil TF-IDF dapat dilihat pada gambar 2. Perhitungan nilai TF secara manual dilakukan dengan persamaan berikut:

$$tfi = .freq_i(d_j) \sum_{i=1}^k freq_i(d_j) \quad (7)$$

Kemudian, dilanjutkan dengan mencari nilai IDF dengan persamaan berikut

$$idf_i = \log |D| / \{d: t_i \in d\} \quad (8)$$

data	jabatan	acara	adli	anies	capres	dukung	ganjar	jakoni	kawan	keadlian	partai	pemilih	persepsi	prabowo	prsiden	rutin	sambung	senang
1	0	0	0.127829	0.145356	0	0.072821	0	0	0.159755	0.155497	0	0	0	0.053897	0	0	0	0
2	0	0.132063	0.022323	0	0.132514	0.267474	0.097381	0.293408	0	0	0.067994	0	0.044285	0.142524	0.075415	0.044285	0	0.032123
3	0.787122	0	0	0	0.137938	0	0	0.080946	0	0.042466	0	0	0.049453	0.044157	0	0	0	0
4	0	0.008452	0	0	0	0	0	0.103719	0	0.163441	0	0	0.095166	0.1133	0	0	0	0
5	0	0	0	0	0.071692	0	0.168075	0	0.132428	0	0	0	0.057623	0	0	0.091801	0	0
6	0	0	0	0	0.124868	0	0.195162	0	0.057663	0	0	0.044767	0.039973	0	0.090254	0	0	0
7	0	0	0	0	0	0	0	0	0	0.022826	0	0	0	0	0	0	0	0
8	0	0	0.151241	0	0.115709	0	0	0	0	0	0	0	0	0	0.021016	0	0	0
9	0	0	0.122236	0	0.141594	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0.032674	0.194865	0.016163	0.130409	0	0	0	0.039808	0	0.129659	0	0.006899	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0.300835	0.036881	0	0	0	0	0	0	0
12	0	0	0	0.059866	0	0	0	0	0	0.05016	0	0	0	0	0	0	0	0
13	0	0.040286	0	0.145875	0	0.031132	0	0	0	0.05016	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0.028648	0.031582	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0.054004	0	0	0	0	0	0	0	0
16	0	0.071356	0	0.110733	0	0	0	0	0	0	0	0	0.068974	0.020529	0	0	0	0
17	0	0	0.093449	0.054776	0.040228	0.053319	0	0.036126	0	0.154867	0.142318	0	0	0	0.039463	0	0	0
18	0	0	0.106452	0	0	0	0	0	0	0.07995	0	0	0	0.103045	0.036948	0	0	0
19	0	0.018874	0	0.023239	0.022722	0.058694	0	0.048709	0.093781	0.02133	0.046999	0	0	0.383125	0.021711	0	0	0.076527
20	0	0.063811	0	0.066017	0	0	0	0	0	0	0	0	0	0.061981	0	0	0	0

Gambar 2. Contoh Hasil TF-IDF

B. Metode Naïve Bayes

1) Confusion Matrix

Tabel 5. Confusion Matrix Metode Naïve Bayes

		Actual Label	
		Non Hoax	Hoax
Predict Label	Non Hoax	27	2
	Hoax	0	31

Tabel 5 menunjukkan simulasi *Naïve Bayes* dengan nilai sebagai berikut

1. Nilai *True Negative* (TN) sebesar 27
2. Nilai *False Positive* (FP) sebesar 2
3. Nilai *False Negative* (FN) sebesar 0
4. Nilai *True Positive* (TP) sebesar 31

2) Evaluasi Kinerja

1. $Accuracy = \frac{31+27}{27+2+0+31} = \frac{58}{60} = 0,96666 = 0,97$ (9)
2. $Precision(p) = \frac{31}{31+2} = \frac{31}{33} = 0,9393939 = 0,94$ (10)
3. $Recall (R) = \frac{31}{31+0} = \frac{31}{31} = 1,00$ (11)
4. $F1 - Score = \frac{2*0,939*1,000}{0,939+1,000} = \frac{1,878}{1,939} = 0,96854 = 0,97$ (12)

Berdasarkan metode *Naïve Bayes* ini maka nilai akurasi dihasilkan sebesar 97%, nilai presisi 94%, nilai recall sebesar 100%, dan nilai *F-1 Score* sebesar 97%.

C. Metode Support Vector Machine

1) Confusion Matrix

Tabel 6. Hasil Confusion matrix Metode SVM

		Actual Label	
		Non Hoax	Hoax
Predict Label	Non Hoax	27	2
	Hoax	1	30

Tabel 6 menunjukkan nilai simulasi SVM sebagai berikut:

1. Nilai *True Negative* (TN) sebesar 27
2. Nilai *False Positive* (FP) sebesar 2
3. Nilai *False Negative* (FN) sebesar 1
4. Nilai *True Positive* (TP) sebesar 30

2) Evaluasi Kinerja

1. $Accuracy = \frac{30+27}{30+2+1+27} = \frac{57}{60} = 0,95$ (13)
2. $Precision(p) = \frac{30}{30+2} = \frac{30}{32} = 0,9375 = 0,94$ (14)
3. $Recall (R) = \frac{30}{30+1} = \frac{30}{31} = 0,96774193548 = 0,97$ (15)
4. $F1 - Score = \frac{2*0,937*0,967}{0,937+0,967} = \frac{1,812}{1,904} = 0,9516806 = 0,95$ (16)

Penggunaan Metode SVM diperoleh nilai akurasi sebesar 95%, nilai presisi sebesar 94%, nilai *recall* sebesar 97%, dan nilai *F-1 Score* sebesar 95%.

Hasil evaluasi pada penelitian ini menunjukkan, hasil akurasi sebesar 97% untuk Metode *Naïve Bayes* dan akurasi sebesar 95% untuk Metode SVM. Hal ini menunjukkan, dengan perlakuan yang sama, Metode *Naïve Bayes* memiliki akurasi yang lebih tinggi dibandingkan dengan Metode SVM. Maka, *Naïve Bayes* dapat dikatakan sebagai metode yang lebih baik sebagai pengklasifikasi teks berita *hoax* dan *non hoax* pada berita atau teks mengenai pemilihan umum presiden di Indonesia tahun 2024.

V. PENUTUP

A. Kesimpulan

Berdasarkan hasil evaluasi kinerja yang telah dijalankan diperoleh kesimpulan yaitu akurasi sebesar 97% diperoleh oleh Metode *Naïve Bayes* dan akurasi sebesar 95% diperoleh Metode SVM. Hal ini menunjukkan, dengan perlakuan yang sama, Metode *Naïve Bayes* memiliki akurasi yang lebih tinggi dibandingkan dengan Metode SVM Maka, *Naïve Bayes* dapat dikatakan sebagai metode yang lebih baik sebagai pengklasifikasi teks berita *hoax* dan *non hoax* pada berita atau teks mengenai pemilihan umum presiden di Indonesia tahun 2024.

B. Saran

Saran ke depan yang dapat dilakukan adalah, peneliti dapat mengkombinasikan metode lainnya untuk melakukan klasifikasi hoaks dan juga menambah data-data dengan berita lainnya yang sedang *trending* atau *viral*.

DAFTAR PUSTAKA

[1] C. A. Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier," *J. Comput. Sci.*

- Inf. Technol.*, vol. 4, no. 1, pp. 207–215, 2024.
- [2] R. Vindua and A. U. Zailani, “Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [3] 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, 2022, [Online]. Available: <https://ojs.utmmataram.ac.id/index.php/explore/article/view/617>
- [4] Y. Findawati, U. Indahyanti, Y. Rahmawati, and R. Puspitasari, “Sentiment Analysis of Potential Presidential Candidates 2024: A Twitter-Based Study,” *Acad. Open*, vol. 8, no. 1, pp. 6–17, 2023, doi: 10.21070/acopen.8.2023.7138.
- [5] C. S. Sriyano and E. B. Setiawan, “Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3396–3405, 2021.
- [6] E. A. Sosiawan and R. Wibowo, “Kontestasi Berita Hoax Pemilu Presiden Tahun 2019 di Media Daring dan Media Sosial,” *J. Ilmu Komun.*, vol. 17, no. 2, p. 133, 2020, doi: 10.31315/jik.v17i2.3695.
- [7] S. Kharya and S. Soni, “Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection,” *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 32–37, 2016, doi: 10.5120/ijca2016908023.
- [8] B. Imran, Zaeniah, Sriasih, S. Erniwati, and Salman, “Data Mining Using a Support Vector Machine, Decision Tree, Logistic Regression and Random Forest for,” *J. INFOKUM*, vol. 10, no. 2, pp. 792–802, 2022.
- [9] E. I. Setiawan, S. Johanes, A. T. Hermawan, and Y. Yamasari, “Deteksi Validitas Berita pada Media Sosial Twitter dengan Algoritma Naive Bayes,” *J. Intell. Syst. Comput.*, vol. 3, no. 2, pp. 55–60, 2021, doi: 10.52985/insyst.v3i2.164.
- [10] F. Rahutomo, I. Y. R. Pratiwi, Pratiwi, and D. M. Ramadhani, “EKSPERIMEN NAIVE BAYES PADA DETEKSI BERITA HOAX BERBAHASA INDONESIA,” *J. Penelit. Komun. dan Opini Publik Vol.*, vol. 23, no. 1, pp. 271–274, 2019, doi: 10.1109/ICAC3N56670.2022.10074374.
- [11] D. Maulina and R. Sagara, “Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency-Inverse Document Frequency,” *J. Mantik Penusa*, vol. 2, no. 1, pp. 35–40, 2018.
- [12] U. Minuril, Alvanof Mulia Mahendra, and Rahmat, “Analisa Dan Deteksi Konten Hoax Pada Media Berita Indonesia Menggunakan Machine Learning,” *J. Teknol. Terap. Sains 4.0*, vol. 1, no. 2, 2020.
- [13] S. Indah Nurhafida and F. Sembiring, “Analisis Text Clustering Masyarakat Di Twitter Mengenai Mcdonald’Sxbts Menggunakan Orange Data Mining,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 28–35, 2021.
- [14] A. Wijaya, C. Rozikin, and B. N. Sari, “Penerapan Text Mining Untuk Klasifikasi Judul Berita Hoax Vaksinasi COVID-19 Menggunakan Algoritma Support Vector Machine,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 11–20, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7058890>
- [15] H. Muhabatin, C. Prabowo, I. Ali, C. L. Rohmat, and D. R. Amalia, “Klasifikasi Berita Hoax Menggunakan Algoritma Naive Bayes Berbasis PSO,” *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 2, p. 156, 2021, doi: 10.51211/itbi.v5i2.1531.
- [16] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- [17] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [18] Heliyanti Susana, “Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet,” *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
- [19] H. Mustofa and A. A. Mahfudh, “Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes,” *Walisongo J. Inf. Technol.*, vol.

- 1, no. 1, p. 1, 2019, doi: 10.21580/wjit.2019.1.1.3915.
- [20] P. K. Laksana Utama, "Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning," *Widya Duta J. Ilm. Ilmu Agama dan Ilmu Sos. Budaya*, vol. 13, no. 1, p. 69, 2018, doi: 10.25078/wd.v13i1.436.
- [21] C. A. Ul Hassan, M. S. Khan, and M. A. Shah, "Comparison of machine learning algorithms in data classification," *ICAC 2018 - 2018 24th IEEE Int. Conf. Autom. Comput. Improv. Product. through Autom. Comput.*, no. January, pp. 1–6, 2018, doi: 10.23919/IConAC.2018.8748995.
- [22] I. P. Putri, "Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 21–28, 2021, doi: 10.33096/ijodas.v2i1.25.