Analisis Sentimen Opini Masyarakat Indonesia Terhadap Konten Deepfake Tokoh Publik

Shane Giorgio Alexander[1], **Amadeus Terra Ananto**[2], **I Putu Adhitya Pratatama Mangku Purnama** [3,] **Bayu Liano Leader Habibullah**[4] **Nur Aini Rakhmawati**[5]
Institut Teknologi Sepuluh Nopember, Surabaya

ARTICLE INFORMATION

Received: September 2023
Revised: Oktobeer 2023
Available online: Oktober 2023

KEYWORDS

Sentiment Analysis, Deepfake, Public Figure, Opinion Mining, Artificial Intelligence

CORRESPONDENCE

Phone:
E-mail:5026211064@mhs.its.ac.id, nur.aini@is.its.ac.id

ABSTRACT

Perkembangan teknologi yang pesat, khususnya Kecerdasan Buatan (AI), telah mendorong transformasi digital yang memberikan efisiensi dan efektivitas dalam berbagai aspek kehidupan. Namun, dampak negatif dari penyalahgunaan teknologi, terutama dalam konteks Deepfake, mengancam privasi, keamanan data pribadi, dan reputasi tokoh publik. Platform media sosial seperti YouTube telah menjadi tempat yang mudah diakses untuk konten Deepfake, sering kali memicu reaksi positif dan negatif dari pengguna. Untuk lebih memahami pandangan masyarakat terhadap konten Deepfake, penambangan teks memainkan peran penting dalam analisis sentimen. Studi ini mengusulkan metode klasifikasi vNaïve Bayes dan pelatihan model IndoBERT untuk menganalisis komentar pengguna YouTube terhadap video Deepfake. Hasil evaluasi menunjukkan tingkat akurasi sebesar 82%, dengan kinerja yang kuat dalam mengklasifikasikan sentimen negatif.

PENDAHULUAN

Perkembangan teknologi yang begitu pesat telah berhasil mempercepat laju digitalisasi dalam berbagai lini kehidupan masyarakat. *Artificial intelligence* adalah salah satu teknologi yang menjadi kunci kesuksesan transformasi digital yang telah berlangsung selama ini. *Artificial Intelligence* atau kecerdasan buatan adalah sistem komputer dengan kemampuan dalam menjalankan tugas-tugas yang umumnya memerlukan kecerdasan manusia melalui proses yang mencakup kemampuan untuk *learning*, *reasoning*, dan *self-correction*, yang begitu mirip dengan cara manusia dalam menganalisis informasi sebelum membuat keputusan [1]. Melalui kecerdasan buatan, berbagai kegiatan sehari-hari masyarakat secara otomatis dapat terselesaikan dengan hasil yang lebih efisien dan efektif. Efektivitas dan efisiensi kecerdasan buatan ini sangat membantu masyarakat dalam mempercepat proses pengambilan keputusan [2].

Kehadiran dari teknologi *Artificial Intellegence* nyatanya tidak selalu memberikan manfaat positif bagi masyarakat, melainkan terdapat dampak negatif yang bersifat merugikan apabila disalahgunakan. Dampak negatif dari teknologi ini berpotensi mengakibatkan beragam masalah seperti pelanggaran privasi pengguna, kerentanan terhadap keamanan data pribadi, dan sejumlah masalah lain yang dapat mempengaruhi masyarakat [3]. Saat ini, penyalahgunaan teknologi *Artificial Intelligence* semakin ramai diperbincangkan, terutama ketika mulai banyak bermunculan konten dari teknologi *Deepfake*. *Deepfake* merupakan produk dari teknologi *Artificial Intelligence* yang menggabungkan, menggantikan, dan menyisipkan gambar serta potongan video untuk membuat konten palsu namun terlihat seperti aslinya [4]. Konten *Deepfake* yang bertebaran di media sosial umumnya bernuansa komedi dan hiburan, namun tidak jarang mengandung unsur negatif seperti pornografi dan berita bohong (*hoaks*). Konten *Deepfake* yang bermuatan negatif ini biasanya menargetkan tokoh-tokoh publik seperti artis, pejabat negara, atau pihak-pihak tertentu yang memegang peranan penting dengan tujuan yaitu merusak reputasi dan mengurangi tingkat kepercayaan masyarakat terhadap tokoh publik tersebut [5]. Hal ini tentu perlu disikapi dengan serius dikarenakan konten *Deepfake* yang begitu sulit untuk dikenali oleh masyarakat awam akhirnya membuat konten ini menjadi mudah untuk menjadi *trending topic*. Popularitas konten *deepfake* akhirnya mengundang perdebatan publik yang sengit, ditandai dengan munculnya berbagai macam komentar dari masyarakat.

Youtube adalah salah satu *platform* media sosial yang menjadi sarang bagi konten-konten bermuatan negatif sehingga sangat mudah diakses dan akhirnya menimbulkan perilaku negatif dari para penggunanya. Konten *Deepfake* juga akhirnya banyak ditemukan pada *platform* ini, salah satunya konten *Deepfake* berupa video yang menampilkan sosok Nagita Slavina,seorang aktris kelas atas yang direkayasa seolah-olah tanpa memakai busana [6]. Kehadiran video ini sontak menggemparkan dunia maya, dan akhirnya memicu beragam tanggapan dari berbagai pihak yang merespons dan tidak sedikit yang mempertanyakan terkait keaslian dari video tersebut. Tingginya angka partisipasi pengguna *Youtube* dalam menyampaikan opininya pada konten *Deepfake* ini berpotensi besar untuk dapat menghasilkan informasi dari polaritas positif dan negatif dalam opini-opini tersebut. Opini tersebut disampaikan melalui kolom komentar *Youtube* dan tersimpan sebagai data yang tidak terstruktur serta mengandung banyak *noise* sehingga diperlukan upaya penambangan teks atau yang dikenal sebagai *teks mining* [7]. *Teks mining* atau eksplorasi teks merupakan proses ekstraksi pola, informasi, dan pengetahuan yang bermanfaat dari sejumlah data teks yang tak terstruktur, dengan tujuan untuk mengidentifikasi pola data, tren, dan ekstraksi pengetahuan potensial dari data teks [8]. *Teknis mining* menjadi langkah utama yang harus dilakukan dalam proses analisis sentimen Analisis sentimen merupakan proses pendekatan dalam mengolah komentar dan umpan balik yang diberikan oleh individu melalui berbagai media terkait dengan suatu produk, jasa, atau entitas tertentu [9].

Banyak penelitian saat ini tengah mengeksplorasi terkait analisis sentimen sebagai topik yang sangat menarik untuk dijadikan subjek jurnal ilmiah. Penelitian sebelumnya yang dilakukan oleh Tanthy Tawaqalia Widowati pada tahun 2020 menganalisis sentimen yang muncul dalam berbagai tweet yang membahas pendapat tentang Nadiem Makarim, yang pada tahun tersebut baru saja diangkat menjadi Menteri Pendidikan dan Kebudayaan di Indonesia. Penelitian ini melibatkan dua metode pengklasifikasian berbeda yakni, *Naive Bayes* dan SVM. Hasil penelitian menunjukkan bahwa dalam penggunaan algoritma *Naive Bayes*, akurasi mencapai 91.48%, sementara SVM mencapai 85.47%. Dalam hal presisi, Naive Bayes mencapai 89.28%, sedangkan SVM mencapai 90.95%. Namun, dalam hal *recall*, *Naive Bayes* mencapai 91.58%, sementara SVM hanya mencapai 76.18% [10]. Penelitian selanjutnya dilakukan oleh Debora Chrisinta dan Justin Eduardo dalam menganalisis sentimen masyarakat terhadap pejabat publik menggunakan data dari Twitter di tahun 2023. Hasil analisis menunjukkan bahwa masyarakat lebih sering

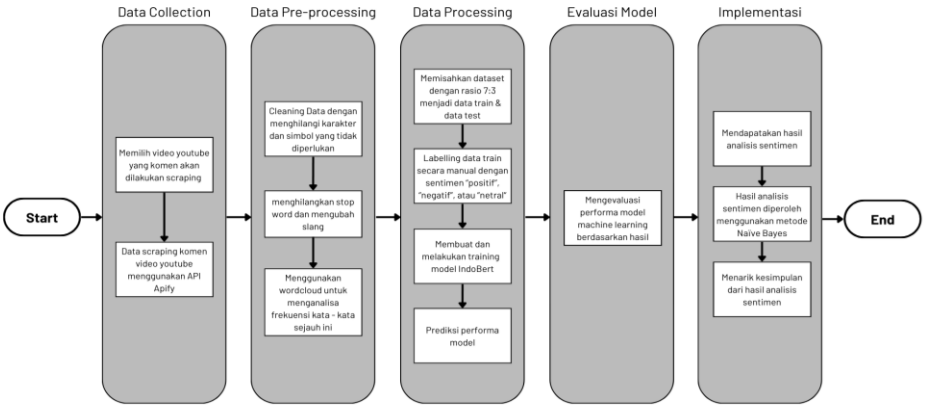
memberikan penilaian negatif. Dalam hal kinerja algoritma, akurasi mencapai 64,55% dengan tingkat kesalahan sebesar 35,45% [11]

Pada penelitian ini, analisis sentimen digunakan untuk mengevaluasi kontribusi konten *deepfake* yang melibatkan tokoh publik dalam mempengaruhi reaksi masyarakat yang dituangkan dalam bentuk komentar, dengan tujuan untuk memahami sejauh mana masyarakat terpancing oleh konten *deepfake*. Berbeda dengan penelitian yang disebutkan di atas, penelitian ini mengembangkan dataset dari kolom komentar *Youtube l*alu mengklasifikasinya ke dalam kategori positif, netral, atau negatif. Untuk melakukan *sentiment analysis* / *opinion mining* terhadap opini masyarakat Indonesia terhadap konten video *deepfake* melibatkan tokoh publik. Sebagian besar algoritma untuk analisis sentimen didasarkan pada pengklasifikasi yang dilatih menggunakan koleksi data teks yang telah dianotasi. Sebelum pelatihan, data diproses terlebih dahulu untuk mengekstraksi fitur-fitur utama [12]. Kemudian, penulis menggunakan metode klasifikasi Naïve Bayes, yang merupakan klasifikasi probabilitas berdasarkan teorema Bayes, dengan pertimbangan asumsi independensi Naïve (kuat). Metode ini diperkenalkan dengan nama yang berbeda ke dalam komunitas pengambilan teks dan tetap menjadi metode yang populer untuk pengkategorian teks, yaitu dengan pengklasifikasian suatu dokumen / kalimat sebagai milik satu kategori atau yang lain dengan frekuensi kata sebagai fitur [13]. Untuk mengklasifikasikan teks atau dokumen ke dalam kategori sentimen tertentu, seperti label "positif", "negatif", atau "netral" untuk melatih model. Pada penelitian kali ini terdapat lima tahapan yaitu *data collection, data pre-processing, data processing,* evaluasi model *sentiment analysis*, dan implementasi model. negatif menggunakan metode *Naive Bayes classifier*.

Tujuan dilakukannya analisis sentimen opini masyarakat Indonesia pada komentar video youtube terhadap konten *deepfake* melibatkan tokoh publik dimana penelitian ini untuk memberikan pemahaman yang lebih baik tentang cara masyarakat Indonesia menanggapi serta merespon penggunaan deepfake khususnya pada tokoh politik untuk menyebarkan hoax. Menggunakan machine learning penulis dapat menganalisa dataset yang berupa komentar - komentar penonton sebanyak 300 komentar lalu menggunakan analisis sentimen penulis akan mendapatkan gambaran keseluruhan mengenai opini dan pandangan masyarakat dalam penggunaan deepfake pada tokoh publik untuk menyebarkan hoax. Melalui pemahaman ini, diharapkan dapat mendorong para pembaca untuk dapat mengembangkan strategi dan pola pikir untuk menghadapi tantangan yang muncul seiring dengan perkembangan teknologi yang dapat disalahgunakan khususnya pada kasus ini yaitu penggunaan deepfake pada tokoh publik untuk menyebarkan berita hoax.

METODE PENELITIAN

Pada penelitian kali ini, penulis memilih untuk menggunakan salah satu paradigma dari machine learning yaitu semi-supervised learning. Dimana semi-supervised learning mengacu pada 'semi-supervised classification', menggunakan data yang sudah dilabel dan data yang tidak berlabel dengan tujuan mengklasifikasi. Semi-supervised learning menggunakan data yang sudah diberikan label untuk training dan data tanpa label testing dimana semi-supervised learning kontras dengan supervised learning dimana semua data diberi label dan juga unsupervised learning dimana semua data tidak diberikan label [14]. Analisis sentimen bertujuan untuk mendeteksi opini-opini termasuk kategori opini positif, negatif, atau netral. Kegunaan dari analisis sentimen di internet adalah untuk mengetahui informasi sentimen publik terkait tokoh politik, film, hotel dan obyek wisata, maskapai penerbangan, dan sebagainya. Hasil dari proses klasifikasi sentimen adalah sekumpulan dokumen yang dapat dipakai untuk menyimpulkan kepuasan pelanggan terkait tokoh politik, layanan atau produk [15].



Gambar 1. Gambaran Umum Tahapan Sentimen Analisis.

Pada Gambar 1. terdapat gambaran umum mengenai jalannya sentimen analisis yang penulislakukan pada penlitian kali ini. Dimana penjelasan tahapan - tahapan sebagai berikut :

1. Data Collection

Data collection atau metode pengumpulan data adalah teknik atau cara-cara yang dapat digunakan oleh penulisuntuk mengumpulkan data. Metode pengumpulan data sebagai suatu metode yang independen terhadap metode analisis data atau bahkan menjadi alat utama metode dan teknik analisis data [16]. Data yang dikumpulkan dalam penelitian akan digunakan untuk menguji hipotesis atau menjawab pertanyaan pada rumusan masalah dan kemudian akan digunakan sebagai dasar dalam pengambilan kesimpulan atau keputusan. Salah satu metode pengumpulan data dalam penelitian yaitu metode *web scraping*.

Tahap pengumpulan data pada penelitian ini dilakukan dengan *web scraping*. Web scraping adalah proses mengambil dokumen yang setengah terstruktur dari internet, khususnya halaman web dengan bahasa markup seperti HTML (*HyperText Markup Language*) atau XHTML (*Extensible HyperText Markup Language*), serta menganalisisnya untuk mengambil data

spesifik dari halaman tersebut guna digunakan untuk tujuan lain [17]. Apify adalah tool *web scrapping* mampu terintegrasi ke berbagai website dan melakukan ekstraksi data menjadi berbagai format, salah satunya Excel [18]. Pengumpulan data Dengan menggunakan Apify untuk data *scraping* komentar video youtube, didapatkan 300 jumlah komentar dari berbagai user dari video - video tersebut.

2. *Data Pre-processing*

Data Pre-processing merupakan langkah pertama dalam menyiapkan data dalam dataset dari *data collection* untuk diproses Tahapan *pre-processing* dilakukan untuk memperbaiki struktur kata yang ada pada tweet sehingga pada tahapan selanjutnya data tersebut mudah dianalisis kata-katanya [19]. Pembersihan data dilakukan pada tahapan ini dengan pertama - tama menghilangkan karakter dan simbol yang tidak diperlukan, menghilangkan *stopword* seperti "di", "ya", "itu", dll yang tidak berkontribusi terhadap sentimen data dan disebut sebagai *noise* [20]. Selanjutnya mengubah kata - kata *slang* dan juga memperbaiki beberapa salah ketikan dengan menggunakan diksi yang sudah umum dipakai, lalu terakhir penulis menghasilkan sebuah *wordcloud* menggunakan data keseluruhan pada tahap pre-processing untuk menganalisa frekuensi kata yang muncul. *Wordcloud* adalah visualisasi dari frekuensi kata dalam dataset, di mana kata-kata yang sering muncul memiliki bentuk visualisasi yang lebih besar visualisasi kata-kata lain [21]. Melalui *wordcloud*, analisis opini masyarakat akan lebih mudah karena bisa melihat poin-poin terpenting saja melalui kata-kata tersebut.

3. *Data Processing*

Pada tahapan ini penulis menggunakan metode *semi-supervised learning* yang merupakan metode yang bercabang dari *machine learning*. *Semi-supervised learning* adalah cabang dari *machine learning* yang berkaitan dengan penggunaan data yang berlabel serta data yang tidak berlabel untuk melakukan beberapa *learning tasks* [22]. Dalam penelitian ini, terdapat tiga jenis sentimen yang digunakan untuk mengklasifikasikan komentar yaitu "positif", "negatif", dan "netral". Training data akan di label secara manual dengan ketentuan :

a. Jika komentar bersentimen baik dimana dalam parameter penulis komentar tersebut menunjukan komentar yang tidak merasa bahwa *Deepfake* ini hal buruk atau *Deepfake* bukan merupakan ancaman, komentar akan diberi label positif dengan angka 1.
b. Jika komentar tidak memiliki konteks terhadap topik atau tidak bersentimen baik/buruk, komentar akan diberi label netral dengan angka 0.
c. Jika komentar bersentimen buruk dimana dari komentar terlihat kekhawatiran dan juga rasa mengkomen bahwa deepfake merupakan hal yang membahayakan, komentar akan diberi label negatif dengan angka 2.

Selanjutnya penulismelatih model *pre-trained* IndoBERT. IndoBERT merupakan *pre-trained* model yang dapat membaca banyak baris data text sekaligus dengan lebih dari 220 juta kata yang dikumpulkan dari tiga sumber utama yaitu Wikipedia Bahasa Indonesia (74 juta kata), artikel berita dari Kompas, Tempo, dan Liputan6 (total 55 juta kata), dan Korpus Web Bahasa Indonesia (90 juta kata) [9]. Oleh karena itu, penulis gunakan untuk melakukan analisis sentimen penulis dengan tujuan utama untuk melakukan proses yang penuli ssebut *tokenizing*. Terakhir, dengan model yang sudah dilatih menggunakan dataset penulis sejauh ini, penuli smelakukan prediksi terhadap performa model yang sudah dihasilkan.

4. Evaluasi Model

Evaluasi model adalah proses menganalisis kinerja model *machine learning* dengan menggunakan berbagai metrik/parameter. Evaluasi model sangat penting dalam menilai efektivitas sebuah model selama fase penelitian awal dan juga berperan dalam pemantauan model kedepannya. Hasil dari algoritma pembelajaran perlu dinilai dan dianalisis dengan benar, sehingga dapat mengevaluasi performa algoritma pembelajaran yang berbeda. Performa klasifikasi diwakili oleh nilai skalar dalam metrik yang berbeda seperti *accuracy*, *sensitivity*, dan *specifity* [23]. Penulis menggunakan metrik evaluasi standar untuk mengukur performa model dengan menggunakan metrik *True Positive*, *True Negative*, *False Positive*, dan *False Negative* dimana Sebuah *true positive* adalah hasil di mana model dengan benar memprediksi kelas positif. Demikian pula, sebuah *true negative* adalah hasil di mana model dengan benar memprediksi kelas negatif. Sebuah *false positive* adalah hasil di mana model dengan salah memprediksi kelas positif. Dan sebuah *false negative* adalah hasil di mana model dengan salah memprediksi kelas negatif. Dengan menggunakan metrik - metrik tersebut penulis dapat mengkalkulasikan akurasi, *recall*, presisi, dan F1-*score* menggunakan persamaan 1, 2, 3, dan 4 [24].

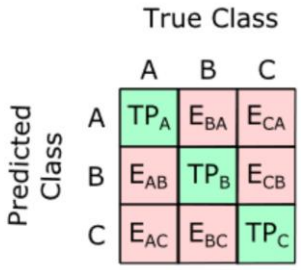
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1 - Score = \frac{2*TP}{2*TP+FP+FN} \quad (4)$$

Pada persamaan (1) yaitu *accuracy* seberapa sering model / mesin membuat prediksi yang benar. Persamaan (2) *recall* menunjukan perbandingan antara *True Positive* (TP) dengan banyaknya data yang sebenarnya positif sedangkan *precision* yang ditunjukan pada persamaan (3) mengukur seberapa banyak prediksi mesin positif yang sebenarnya positif. Terakhir, F1 *score* yang ditunjukan pada persamaan (4) menggabungkan *precision* dan *recall* dalam satu metrik, F1 *score* sangat penting untuk menentukan performa model *machine learning*.



Gambar 2. *Confusion matrix* untuk mengklasifikasi multi-kelas. [10, Gambar. 3].

Tahap evaluasi ini dilakukan dengan menggunakan *confusion matrix* yang menguji kinerja mesin / model untuk menunjukkan detail klasifikasi yang benar dan salah untuk setiap kategori (data uji serta data prediksi) dengan lebih rinci. *Confusion matrix* ini berisi semua informasi yang belum diproses tentang prediksi yang dilakukan oleh model klasifikasi pada kumpulan data tertentu. Indikator kinerja dapat berupa, misalnya, presisi, *recall*, atau skor F [18]. Gambar 2 menunjukan *confusion matrix* untuk masalah klasifikasi multi-kelas dengan tiga kelas (A, B, dan C) [25]. Tiap *cell* yang berwarna hijau merepresentasikan jumlah prediksi benar sedangkan sel berwarna pink mengindikasikan jumlah prediksi salah yang dihasilkan oleh model. Ketika sampel positif diklasifikasikan dengan kelas positif maka prediksi tersebut *true positive* (TP). Ketika sampel positif diklasifikasikan dengan kelas negatif maka prediksi tersebut *false negative* (FN) atau disebut Type II error. Apabila sampel negatif diklasifikasikan dengan kelas positif maka prediksi tersebut *false positive* (FP) yang merupakan *false alarm* atau Type I error. Ketika sampel negatif diklasifikasikan dengan kelas negatif maka prediksi tersebut *true negative* (TN) [26].

5. Implementasi Model

Setelah model dinilai sudah memiliki evaluasi yang cukup, model diimplementasikan pada dataset lebih besar. Pada tahap ini, data yang terkumpul mulai dilabeli sebagai positif, negatif, dan netral oleh model. Setelah pemberian label selesai, model akan membuat kolom baru pada dataset berdasarkan model.

HASIL DAN PEMBAHASAN

Bagian ini akan membahas mengenai hasil analisis sentimen dengan dataset yang diperoleh dari data scraping pada komentar dari beberapa video youtube channel Indonesia yang menurut penulis relevan dengan penelitan pada dasarnya yang mengandung keyword *Deepfake* dan *Hoax* serta mengandung tokoh publik sebagai bahan *deepfake*.

3.1. Pembersihan dan Pemrosesan Data

Dataset yang diperoleh berjumlah 300 data mentah yang mempunyai *header* yaitu *author* sebagai penulis dari komentar, *cid* yaitu id untuk komentar, *comment* yang berisikan komentar yang tertulis, *commentsCount* yang menjelaskan jumlah komentar pada video yang terdapat komentar tersebut, pageUrl untuk URL video youtube bersangkutan, *title* yang menandakan judul video yang dilakukan *data scraping*, dan videoId yaitu id video youtube tersebut. Setelah itu *data pre-processing* dilakukan untuk merestrukturisasi data dan membersihkannya.

Untuk bisa melakukan *sentiment analysis*, penulis membagi *dataset* yang sudah ada dengan rasio 3:7 untuk *training dataset* dan *test dataset* yang membuat *training dataset* memiliki 205 komentar, *training dataset* dengan 143 *training samples* dan 62 *validation samples*, lalu *test dataset* yang memiliki 95 komentar. Setelah mendapatkan 205 komentar sebagai *training dataset* penulis secara manual memberikan label berdasarkan parameter yang telah disebutkan pada bagian *3. Data Processing*.

Tabel 1. Contoh Komentar dengan Sentimen Netral setelah *Pre-Processing*

Comments	**Sentimen**
terimakasih bang ilmunya bermanfaat semangat ijin share	Netral
deepfake emang pisau bermata dua tergantung lihat sisi mana jadi manusia jaman berkembang susah jangan mudah dipercaya	Netral
wowkeren bromanthap ilmunya	Netral

Tabel 2. Contoh Komentar dengan Sentimen Positif setelah *Pre-Processing*

Comments	**Sentimen**
deepfake nya msih kasar tuh klihtan banget palsu	Positif
kalau ga pernah jarang upload foto internet medsos insyaallah aman saja	Positif
deepfake sekarang kelihatan palsu	Positif

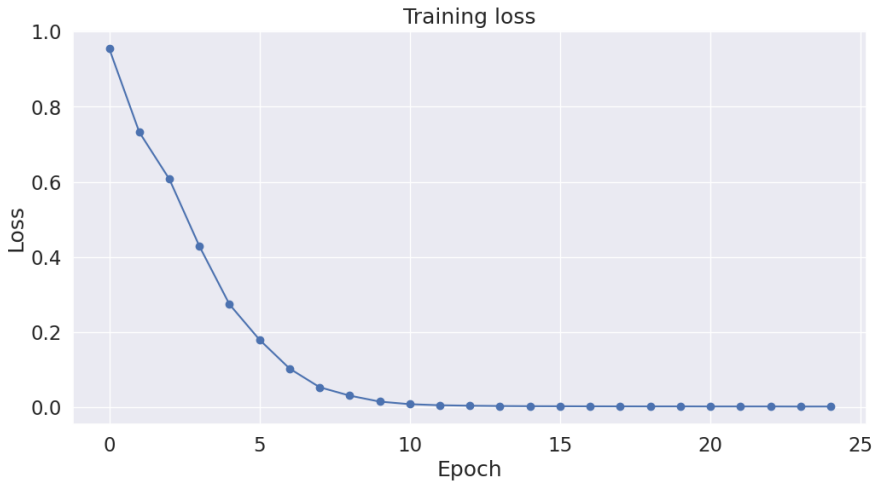
Tabel 3. Contoh Komentar dengan Sentimen Negatif setelah *Pre-Processing*

Comments	Sentimen
semakin real semakin menyeramkan di salahgunakan orang terutama buat kriminal penipuan harusnya aplikasi kaya gini janganlah gampang digunakan semua orang	Negatif
gilaaa parah susah recognice org ini via net kedepannya videocall nyamar jadi president parah tuh teknologi kubur saja soalnya bakal bahaya siapa yang setuju like	Negatif
deepfake sebenarnya jahat jahat orang yang menyalahgunakan teknologi tersebut hal yang senonoh negatif	Negatif

Setelah melakukan pelabelan, akan dilakukan *tokenizing* dengan IndoBERT dengan tujuan untuk mengubah teks atau dokumen menjadi sejumlah token berukuran kecil atau kata-kata yang lebih mudah dikelola. Lalu akan dilakukannya *training* terhadap model *machine learning*, proses pelatihan model *machine learning* untuk analisis sentimen melibatkan pembelajaran model tentang fitur-fitur yang ada dalam teks/token dan bagaimana fitur-fitur tersebut berkaitan dengan sentimen tertentu.

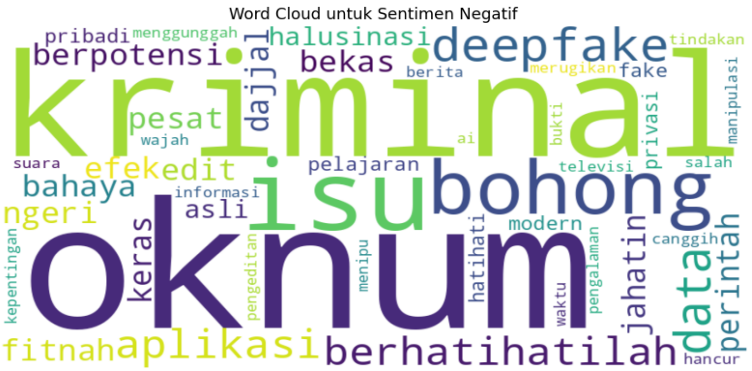
3.2. Analisis Sentimen dan Hasil Algoritma

Mesin dilatih untuk mengklasifikasikan teks atau data input menjadi berbagai kategori sentimen, seperti positif, negatif, atau netral. Proses pelatihan pada penelitian kali ini menggunakan jumlah *epoch* 25 atau 25 iterasi pelatihan mesin dimana model *machine learning* telah menunjukan tingkat kinerja yang stabil/konvergen. Setelah proses pelatihan selesai, mesin dapat digunakan untuk memprediksi atau mengklasifikasikan sentimen dari data yang belum pernah dilihat sebelumnya.



Gambar 6. Diagram *Training Loss* dengan 25 Iterasi

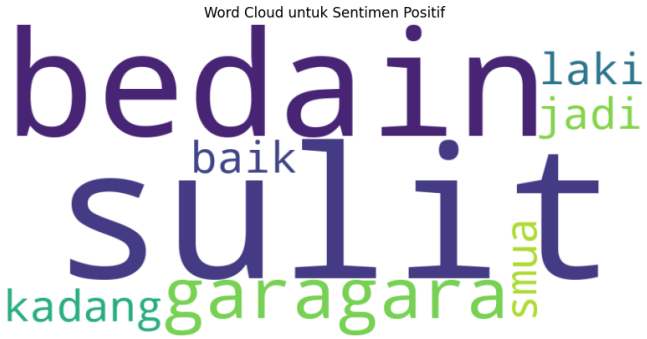
Untuk menganalisa frekuensi dan sebaran kata - kata untuk sentimen negatif dan positif sebagai bahan pertimbangan juga dalam menganalisa hasil sentimen analysis penulis menggunakan *wordcloud*. *Wordcloud* akan menampilkan kata - kata dengan frekuensi muncul terbanyak pada masing - masing sentimen positif dan negatif dimana ukurang *font size* yang besar menunjukan kata tersebut muncul lebih banyak dibandingkan dengan kata yang memiliki *font size* lebih kecil.



Gambar 7. *Wordcloud* untuk Token Sentimen Negatif *Test Data*

Wordcloud pada gambar 7. menunjukan *wordcloud* kata - kata pada test data yang dilabeli model sebagai data bersentimen negatif. Kata - kata yang memiliki frekuensi tertinggi untuk data bersentimen negatif yang berjumlah 22 data yang meliputi kata -

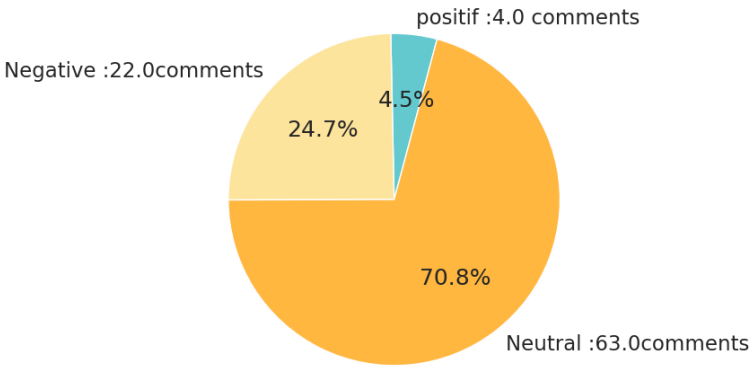
kata seperti kriminal, oknum, isu, berhati-hatilah, ngeri, dan bohong. Kata - kata tersebut menunjukan pandangan masyarakat terhadap penggunaan *deepfake* pada tokoh publik untuk menyebarkan *hoax* dimana masyarakat masih merasakan rasa khawatir terhadap teknologi *deepfake* dan juga menganggap *deepfake* sebagai teknologi yang membahayakan.



Gambar 8. *Wordcloud* untuk Token Sentimen Positif Test Data

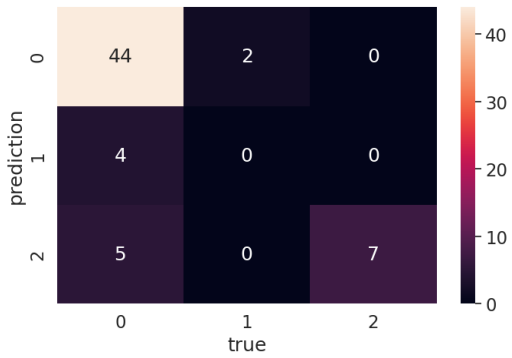
Pada gambar 8. *wordcloud* menunjukan kata - kata yang lebih sedikit, ini dapat dilihat dikarenakan data yang dilabeli model bersentimen positif berjumlah hanya 4 dengan presisi serta *recall* yang rendah. Kata - kata yang memiliki frekuensi tinggi pada *wordcloud* diatas seperti kata bedain, sulit, baik, dan lain - lain. Kata bedain dan sulit memiliki frekuensi yang tinggi dikarenakan pada pelabelan *data training*, komentar seperti "kelihatan banget bedanya" dan semacamnya dilabeli penulis dengan sentimen positif dengan parameter seperti pada bagian *data processing*.

Frekuensi kata-kata yang sering muncul sangat penting untuk digunakan sebagai bobot dalam analisis sentimen. Secara keseluruhan, mesin menilai bahwa dari 95 data untuk dataset *test data* terdapat 4 komentar yang dinilai memiliki sentimen positif, 22 komentar bersifat negatif, 63 komentar bersifat netral, dan sisanya memiliki nilai NaN.



Gambar 9. *Pie Chart* Persentase Pelabelan Sentimen *Test Data*

3.3. Evaluasi



Gambar 10. *Confusion Matrix Test Data*

Dari *confusion matrix test data* dimana angka 0 melambangkan netral, 1 melambangkan positif, dan 2 melambangkan negatif diketahui bahwa prediksi komentar netral dengan aktual netral sebanyak 44 komentar (*true neutral*), prediksi komentar positif dengan aktual positif sebanyak 0 komentar (*true positive*), prediksi komentar negatif dengan aktual negatif sebanyak 7 komentar (*true negative*), komentar aktual netral dengan prediksi positif sebanyak 4 komentar dan prediksi negatif sebanyak 5 komentar, dan prediksi komentar netral dengan aktual positif sebanyak 2 komentar.

Tabel 4. Indikator Kinerja Test Data

	Accuracy	*Precision*	*Recall*	*F1-Score*
0 (Netral)	0.82	0.83	0.96	0.89
1 (Positif)		0.00	0.00	0.00
2 (Negatif)		1	0.58	0.74

Hasil evaluasi model menunjukkan performa yang cukup baik dalam mengklasifikasikan data dengan tingkat akurasi sebesar 82%. Model memiliki performa cukup baik dalam mengidentifikasi kategori negatif dengan tingkat presisi sebesar 83% dan recall sebesar 96%, menghasilkan nilai F1-score sebesar 0.89. Namun, performa model dalam mengklasifikasikan kategori 1 kurang memuaskan dengan presisi dan recall sebesar 0%, sehingga nilai F1-score-nya adalah 0. Model juga memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan kategori 2, dengan presisi 100% dan recall 58%, menghasilkan nilai F1-score sebesar 0.74. Meskipun secara umum model memiliki tingkat akurasi yang baik, perbaikan mungkin diperlukan pada kategori 1 untuk meningkatkan kinerja keseluruhan model dalam mengklasifikasikan data dengan lebih seimbang. Selain itu, nilai rata-rata tertimbang (weighted avg) untuk presisi, recall, dan F1-score menunjukkan bahwa model ini memiliki performa keseluruhan yang baik.

KESIMPULAN

Penelitian ini menginvestigasi pandangan masyarakat terhadap konten *deepfake* yang melibatkan tokoh publik dan mengukur besarnya dampak yang ditimbulkan akibat dari penyebaran konten *deepfake* di media sosial terkhususnya *Youtube*. Proses yang berlangsung dalam penelitian ini adalah proses klasifikasi sentimen terhadap opini yang disampaikan oleh masyarakat Indonesia melalui kolom komentar pada video *Youtube* yang merupakan konten *deepfake*. *Dataset* yang digunakan dalam penelitian ini terdiri dari 300 data yang terbagi ke dalam X data *training* dan Y data *testing*. Proses klasifikasi sentimen dilakukan dengan memanfaatkan model IndoBERT, yang kemudian diklasifikasikan melalui metode s*emi-supervised learning*. Selain itu, hasil prediksi sentimen juga diperoleh melalui penerapan metode *Naive Bayes Classifier*.

Hasil akhir dari penelitian ini telah menunjukkan keberhasilan dalam proses klasifikasi sentimen data komentar pada video *Youtube* dengan kata kunci "deepfake" dan "hoax" yang melibatkan algoritma *Naive Bayes*. Hasilnya menunjukkan bahwa model memiliki akurasi sebesar 82%, mean F1 sebesar 54.3%, recall sebesar 51.34%, dan precision sebesar 61%. Model mengidentifikasi beberapa kata yang memiliki frekuensi tinggi dalam kalimat-kalimat dengan sentimen positif dan negatif. Sentimen negatif tercermin dalam kata-kata seperti "teknologi," "oknum," "bohong," "kriminal," dan lainnya. Sementara itu, sentimen positif tercermin dalam kata-kata seperti "iya," "sini," "bedain," "sulit," "siapa," dan lainnya. Berdasarkan data tersebut, dapat disimpulkan bahwa masyarakat Indonesia cenderung lebih banyak mengungkapkan pendapat negatif daripada positif terkait penggunaan teknologi *deepfake* yang melibatkan tokoh publik. Berdasarkan hasil pengujian terhadap 95 data, 22 data mengindikasikan sentimen negatif, sementara hanya 8 data yang menunjukkan sentimen positif. Meskipun begitu, mayoritas data, yaitu sebanyak 63 data, menunjukkan sentimen netral. Hal ini mengindikasikan bahwa sebagian besar masyarakat Indonesia tampaknya tidak terlalu dipengaruhi dalam menyampaikan pendapat mereka terkait penggunaan deepfake yang melibatkan tokoh publik.

DAFTAR PUSTAKA

[1] N. E. Maaliki and E. Soponyono, "Kebijakan Hukum Pidana Dalam Menanggulangi Tindak Pidana Berita Bohong," *J. Pembang. Huk. Indones.*, vol. 3, no. 1, pp. 59–69, 2021, doi: 10.14710/jphi.v3i1.59-69.

[2] Sipayung, E. M., Maharani, H., & Zefanya, I., "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier", 2016. Tersedia: https://doi.org/10.36706/jsi.v8i1.3250.

[3] T. S. Ramli and R. F. Mayana, "Pemanfaatan Artificial Intelligence Pada Fitur PayLater Aplikasi Shopee Dalam Bidang E-Commerce Dikaitkan Dengan Data Pribadi Konsumen Berdasarkan Hukum Positif Indonesia," vol. 03, no. 04, pp. 1538–1551, 2023, doi: 10.59141/comserva.v3i4.902.

[4] Widodo, A. O., Septiadi, F., & Rakhmawati, N. A. (2023). ANALISIS TREN KONTEN PADA VTUBER INDONESIA MENGGUNAKAN LATENT DIRICHLET ALLOCATION. Jurnal Informatika Dan Rekayasa Elektronik, 6(1), 56–63. https://doi.org/10.36595/jire.v6i1.718

[5] D. Dari, H. Pidana, I. Hukum, F. Ilmu, and U. N. Surabaya, "Jerat hukum penyalahgunaan aplikasi," 2016.

[6] M. F. Hafiz, "Heboh Video Syur Mirip Nagita Slavina polisi pastikan video Hasil Editan," Pikiran Rakyat Mataram, https://mataram.pikiran-rakyat.com/seni-budaya/pr-2223491784/heboh-video-syur-mirip-nagita-slavina-polisi-pastikan-video-hasil-editan (accessed Oct. 3, 2023).

[7] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.

[8] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems*. New Jersey, New Jersey: Pearson Education, Inc., 2011.

[9] Koto, F., Rahimi, A., Lau, J. H. and Baldwin, T. "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained," in Proceedings of the 28th International Conference on Computational Linguistics.Barcelona, Spain. International Committee on Computational Linguistics. 2020. pp. 757-770.

[10] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.

[11] D. Chrisinta and J. E. Simarmata, "Analisis Sentimen Penilaian Masyarakat Terhadap Pejabat Publiks Menggunakan Algoritma Naïve Bayes Classifier," *Komputika J. Sist. Komput.*, vol. 12, no. 1, pp. 93–101, 2023, doi: 10.34010/komputika.v12i1.9638.
[12] P. Gamallo and M. Garcia, "Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets," *8th Int. Work. Semant. Eval. SemEval 2014 - co-located with 25th Int. Conf. Comput. Linguist. COLING 2014, Proc.*, no. SemEval, pp. 171–175, 2014, doi: 10.3115/v1/s14-2026.
[13] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016, doi: 10.5815/ijieeb.2016.04.07.
[14] X. Zhu, "Tr1530." p. 39, 2005, [Online]. Available: http://digital.library.wisc.edu/1793/60444.
[15] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022.
[16] M. Makbul, *Metode Pengumpulan Data Dan instrumen Penelitian*, pp. 1–38, 2021. doi:10.31219/osf.io/svu73
[17] D. F. Setiawan, T. Tristiyanto, and A. Hijriani, "Aplikasi Web Scraping Deskripsi Produk," *J. Teknoinfo*, vol. 14, no. 1, p. 41, 2020, doi: 10.33365/jti.v14i1.498.
[18] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann. Math. Artif. Intell.*, vol. 81, no. 3–4, pp. 429–450, 2017, doi: 10.1007/s10472-017-9564-8.
[19] N. A. Rakhmawati, M. I. Aditama, R. I. Pratama, and K. H. U. Wiwaha, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," J. Inf. Eng. Educ. Technol., vol. 4, no. 2, pp. 90–92, 2020, doi: 10.26740/jieet.v4n2.p90-92
[20] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC - Trends Anal. Chem.*, vol. 132, p. 116045, 2020, doi: 10.1016/j.trac.2020.116045.
[21] T. M. Fahrudin, A. R. F. Sari, A. Lisanthoni, and A. A. D. Lestari, "Analisis Speech-To-Text Pada Video Mengandung Kata Kasar Dan Ujaran Kebencian Dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens Dan Visualisasi Word Cloud," *Skanika*, vol. 5, no. 2, pp. 190–202, 2022, doi: 10.36080/skanika.v5i2.2942.
[22] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020, doi: 10.1007/s10994-019-05855-6.
[23] M. I. Amal, E. S. Rahmasita, E. Suryaputra, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 645–660, 2022, doi: 10.28932/jutisi.v8i3.5483.
[24] A. Yazdinejad, B. Zolfaghari, A. Dehghantanha, H. Karimipour, G. Srivastava, and R. M. Parizi, "Accurate threat hunting in industrial internet of things edge devices," *Digit. Commun. Networks*, no. April 2021, 2023, doi: 10.1016/j.dcan.2022.09.010.
[25] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.
[26] N. I. Wibowo, T. A. Maulana, H. Muhammad, and N. A. Rakhmawati, "Perbandingan Algoritma Klasifikasi Sentimen Twitter Terhadap Insiden Kebocoran Data Tokopedia," JISKA (Jurnal Inform. Sunan Kalijaga), vol. 6, no. 2, pp. 120–129, 2021, doi: 10.14421/jiska.2021.6.2.120-129 [Data set]. Zenodo. http://doi.org/10.5281/zenodo.8401316