



## ANALISIS SENTIMEN PUBLIK TERHADAP PENGGUNAAN TEKNOLOGI DEEFAKE PADA ERA DIGITAL

Arumi Adhiatmico Setiawan <sup>1)\*</sup>, Trista Avrilia Santoso <sup>2)</sup>, Gabriella Jessica Prawira <sup>3)</sup>, Nur Aini Rakhmawati <sup>4)</sup>

Institut Teknologi Sepuluh Nopember <sup>1),2),3),4)</sup>

[5026221007@student.its.ac.id](mailto:5026221007@student.its.ac.id) <sup>1)\*</sup>, [5026221008@student.its.ac.id](mailto:5026221008@student.its.ac.id) <sup>2)</sup>, [5026221175@student.its.ac.id](mailto:5026221175@student.its.ac.id) <sup>3)</sup>, [nur.aini@its.ac.id](mailto:nur.aini@its.ac.id) <sup>4)</sup>

### ABSTRAK

*Deepfake* merupakan sebuah teknologi berbasis kecerdasan buatan (AI), yang memungkinkan untuk melakukan manipulasi terhadap gambar, video, dan audio dengan tingkat realisme yang tinggi. Meskipun teknologi ini termasuk ke dalam salah satu inovasi, namun masih seringkali disalahgunakan untuk menghasilkan konten yang bersifat *hoax*. Penyalahgunaan yang dilakukan seperti untuk melakukan penipuan, memanipulasi, dan mengancam individu maupun sebuah organisasi. Hal tersebut berpotensi dalam merusak reputasi korban dan memengaruhi sentimen dari publik. Penelitian ini bertujuan dalam menganalisis sentimen publik terhadap penggunaan teknologi *deepfake* melalui *platform X*. *Platform X* merupakan sebuah *platform* media sosial yang menggantikan sosial media *twitter*. Analisis sentimen ini dilakukan untuk mengetahui opini publik terhadap topik *deepfake*. Sebanyak 312 data komentar dikumpulkan dengan menggunakan metode *crawling* melalui *platform X* dan diproses menggunakan metode *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen menjadi kategori positif, netral, negatif. Evaluasi dari kinerja model ini menunjukkan tingkat akurasi sebesar 98% yang dikategorikan sebagai sangat baik dengan data 25 komentar negatif, 14 komentar positif, dan 267 komentar netral. Walaupun demikian, keterbatasan data dan perbedaan persepsi dalam *labeling* dapat memengaruhi hasil. Penelitian ini memberikan kontribusi dalam memahami persepsi publik terhadap *deepfake* dan diharapkan dapat digunakan sebagai acuan dalam melakukan pengolahan informasi mengenai *deepfake* kedepannya dengan hasil yang lebih baik.

**Kata Kunci:** *Deepfake*, Kecerdasan Buatan, Analisis Sentimen, SVM

### 1. PENDAHULUAN

Kecerdasan buatan atau yang biasa dikenal dengan *artificial intelligence* atau AI, merupakan sebuah teknologi yang banyak digunakan belakangan ini. AI adalah sebuah teknologi yang dirancang agar sebuah sistem komputer dapat meniru kemampuan intelektual dari manusia. AI juga memungkinkan agar sebuah sistem belajar dari pengalaman, melakukan identifikasi pada pola, membuat dan mengambil keputusan, dan menyelesaikan tugas kompleks dengan cepat [7]. Salah satu bentuk dari kehadiran AI adalah *deepfake*.

*Deepfake* sendiri merupakan sebuah AI yang dapat mengubah piksel pada gambar, yang dimana hasil modifikasinya akan berbeda dengan gambar asli. Selain itu, dengan menggunakan *deepfake*, dapat mengubah tampilan wajah yang dimana hal ini banyak disalahgunakan oleh orang yang tidak bertanggungjawab, seperti dengan membuat video, gambar, maupun suara *hoax*. Padahal, orang yang menjadi korban tidak melakukannya. Fakta tersebutlah yang menjadi kecemasan publik, terutama bagi mereka yang jarang menggunakan media sosial. Terdapat beberapa reaksi yang dialami oleh korban, seperti gangguan *stress* karena reputasi buruk yang dialaminya. Masyarakat saat ini jarang sekali mengolah informasi yang didapatkannya dan hanya diterima mentah-mentah tanpa mengetahui kebenarannya walaupun dari sumber yang tidak kredibel [17].

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap penggunaan *deepfake* yang belakangan ini marak terjadi. Platform yang digunakan adalah Platform X. Penelitian ini akan menganalisis *tweet* dan komentar-komentar mengenai *deepfake* dengan fokus pada bagaimana *tweet* maupun komentar tersebut memengaruhi sentimen positif, negatif, ataupun netral. Dalam penelitian ini, akan digunakan pendekatan analisis *Support Vector Machine* (SVM) untuk menganalisis sentimen dari para pengguna X. Metode SVM digunakan dalam penelitian karena algoritma SVM adalah algoritma dengan kemampuan terbaik, dilihat dari nilai *f1-score* yang memiliki nilai tertinggi dibandingkan dengan algoritma lainnya [2].

Beberapa penelitian sebelumnya telah melakukan eksplorasi terhadap analisis sentimen di media sosial. Namun, penelitian yang membahas mengenai analisis sentimen terhadap penggunaan teknologi *deepfake* di media sosial dengan menggunakan metode *Support Vector Machine* (SVM) masih terbatas. Hal ini berdasarkan beberapa penelitian terdahulu seperti Analisis Sentimen Netizen Terhadap Personal Branding Elon Musk Pada Platform X Dengan Pendekatan Analisis *Support Vector Machine*, Analisis Sentimen Masyarakat Terhadap Undang-Undang Perlindungan Data Pribadi Pada Aplikasi X Dengan Metode *Support Vector Machine* [6], dan penelitian lainnya. Penelitian ini akan menambahkan pemahaman baru mengenai bagaimana teknologi *deepfake* dapat memengaruhi pandangan publik terhadap kemajuan teknologi. Hasil penelitian ini juga dapat memberikan manfaat bagi masyarakat dalam mengelola kebenaran informasi yang diterima.

## 2. TINJAUAN PUSTAKA

### *Deepfake*

*Deepfake* merupakan produk dari kecerdasan buatan yang berbentuk video, gambar, audio, maupun teks sintesis tertentu di mana seseorang dalam audio, gambar, atau video ditukar dengan kemiripan orang lain. Istilah “deepfake” pertama kali diciptakan pada akhir tahun 2017 oleh pengguna layanan *Reddit*. Selanjutnya, Pengguna ini membuat ruang di situs berita dan agregasi online, yang mereka jadikan sebagai tempat berbagi video porno yang menggunakan teknologi pertukaran wajah dari internet [11]. Istilah ini kemudian diperluas hingga mencakup “aplikasi media sintetik” yang sudah ada sebelum laman *Reddit* dan kreasi baru seperti “StyleGAN” – “gambar diam yang tampak realistis dari orang-orang yang tidak ada. Dalam contoh yang lebih baru, *deepfake* dapat berupa suara yang terdengar seperti seseorang yang dikenal. Sebagai contoh, dapat dilihat fenomena yang terjadi pada Mark Zuckerberg dari Facebook dalam video yang sudah diedit yang menggembar-gemborkan betapa hebatnya memiliki miliaran data orang atau perdana menteri Belgia yang mengaitkan virus corona dan pandemi hingga perubahan iklim selama rekaman pidato yang dimanipulasi [10]. Secara garis besar, *Deepfake* dapat digunakan untuk memanipulasi dan mengancam individu dan perusahaan. Namun dengan pemahaman yang lebih baik tentang teknologi, para eksekutif dapat mengambil langkah-langkah untuk melindungi diri mereka sendiri dan perusahaan mereka [4].

### *Platform X*

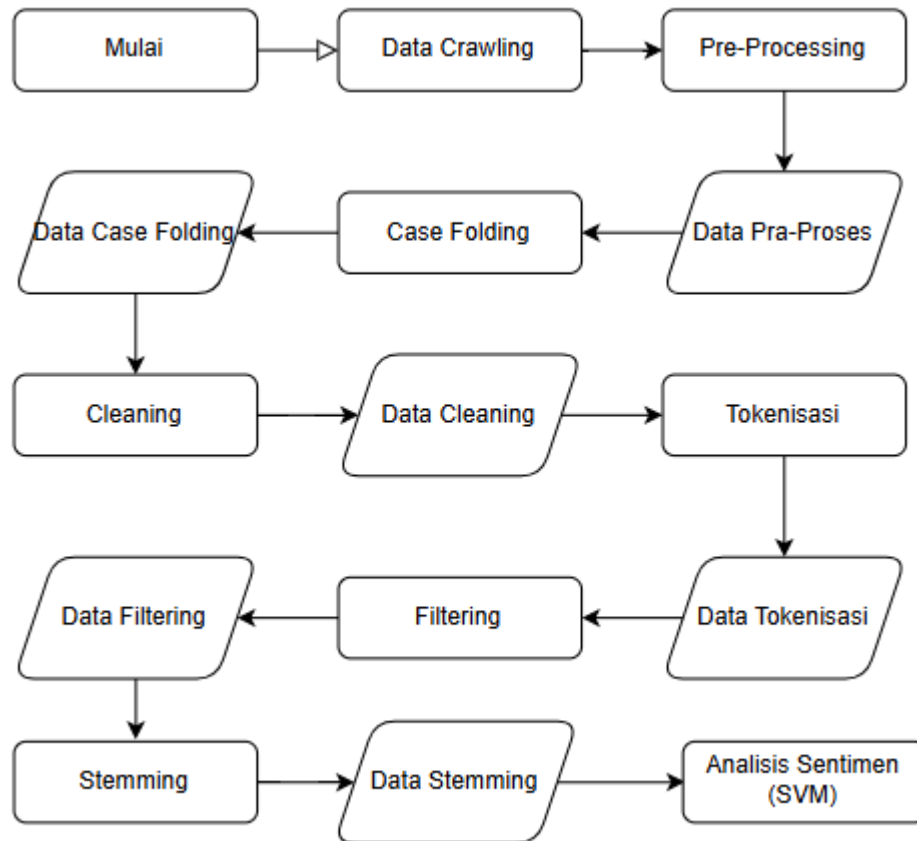
*X* merupakan sebuah *platform* yang menggantikan media sosial *twitter*. Pada Oktober 2022, Elon Musk membeli dan mengganti nama *twitter* dengan harga US\$44 miliar atau setara dengan Rp683 triliun. Elon Musk berencana untuk menjadikan media sosial ini menjadi “wadah yang dapat menyalurkan segalanya.” Elon Musk juga mengatakan bahwa dia memilih *X* karena huruf ini “mewujudkan ketidaksempurnaan di dalam diri kita yang membuat kita menjadi unik.” [16]. Pada *platform* ini, pengguna dapat melakukan unggah komentar, *tweet*, dan gambar, menyukai *postingan*, dan mengutip *posting*, serta mengirim pesan secara langsung melalui DM kepada pengguna lainnya. Pengguna dapat berinteraksi melalui *X* dengan menggunakan gawai, *web browser*, atau secara terprogram melalui *application programming interface* (API) [13].

### *Analisis Sentimen*

Analisis sentimen adalah teknik data tekstual yang menggunakan *Natural Language Processing* (NLP) dan *Machine Learning* (ML) untuk mengevaluasi teks secara otomatis untuk perasaan yang dirasakan oleh penulis (positif, negatif, dan netral) [5]. Tujuan utamanya untuk menganalisis opini, sentimen, penilaian, atau perasaan seseorang terhadap suatu topik, produk, layanan, organisasi, individu, atau kegiatan tertentu. Teknik ini akan mengelompokkan teks yang dianalisis ke dalam konteks positif, negatif, atau netral. Sehingga penilaian dan pandangan masyarakat tentang objek yang dianalisis tersebut akan nampak pada hasil analisis [1].

## 3. METODE PENELITIAN

Pada bab ini akan dipaparkan mengenai metodologi yang digunakan dalam mengolah data *X*. Metodologi tersebut terdiri dari tahapan-tahapan yang dapat dilihat pada *flowchart* sebagai berikut.



Gambar 1 *Flowchart Tahapan*

Pada analisis ini, kami menggunakan Google Collaboratory sebagai tools dalam melakukan pengolahan data, yang merupakan sebuah *platform Machine Learning* dari Google, layanan *online* seperti Jupyter Notebook yang tidak memerlukan pengaturan untuk digunakan dan menyediakan akses gratis ke sumber daya komputasi, termasuk GPU dan TPU. Melalui platform ini, kami menggunakan kode *Python* untuk mengelola data.

### **Data Crawling**

*Data crawling* merupakan salah satu metode yang dilakukan dalam mengumpulkan data. Metode ini dilakukan untuk mengumpulkan informasi penting yang dibutuhkan yang berasal dari *platform X*. Data yang diambil berupa komentar dari beberapa postingan terkait dengan *deepfake* di *platform X*. Dalam melakukan *Crawling* ini, kami memanfaatkan tools dan modul *node.js* dan *tweet-harvest* untuk memudahkan pengambilan data [14].

### **Pra-Proses**

Pra-proses merupakan tahap awal dalam mengolah data yang diterapkan pada *database* untuk menghapus data yang tidak diperlukan, seperti, *noise*, *missing value*, dan data yang tidak konsisten. Tahap ini bertujuan dalam melengkapi *database* hingga konsisten [4]. Agar data yang dihasilkan lebih akurat dan konsisten dan memudahkan dalam melakukan analisis, terdapat beberapa tahapan dalam melakukan Pra-Proses data.

#### **1. Case Folding**

Tahap *case folding* merupakan salah satu tahap dalam pra-proses data. Pada tahap ini, dilakukan untuk menyamakan bentuk ataupun karakter pada data [12]. Karakter yang memiliki huruf kapital seperti ‘A-Z’ akan diubah menjadi huruf kecil seperti ‘a-z’. Selain itu, dapat menghilangkan tanda baca seperti koma untuk memudahkan pembacaan data. Juga, menghapus ruang kosong pada setiap data jika ada.

#### **2. Cleaning**

Tahap *cleaning* merupakan tahap pembersihan data, seperti menghilangkan baris yang kosong atau bernilai *null* dan menghapus baris yang tidak mengandung angka atau huruf [12].

#### **3. Tokenizing**

Tahap *tokenizing* merupakan tahap dimana mengubah setiap kalimat pada data menjadi kata-kata. Contohnya adalah seperti “kasian pak deh” menjadi “kasian”, “pak”, “deh” [8].

#### 4. Stemming

Tahap *stemming* merupakan tahapan untuk mengubah kata yang memiliki imbuhan menjadi kata dasar. Bertujuan untuk mengurangi *unique word* yang harus diolah oleh model [8].

#### 5. Remove Stopword

Tahap *remove stopwords* merupakan tahapan untuk menghapus kata-kata yang tidak berkaitan dengan konteks atau kata-kata yang tidak memiliki kontribusi dalam analisis yang dilakukan. Kata-katanya biasanya didapatkan dari daftar *stopword* [12].

### Analisis Sentimen (SVM)

Untuk melakukan analisis sentimen menggunakan *support vector machine*, dilakukan tahapan-tahapan dalam melakukannya. Berikut adalah tahapan yang dilalui.

#### 1. Polarity & Subjectivity

Tahap *polarity* merupakan tahapan dimana berfungsi dalam melihat kecondongan sentimen dari sebuah teks. Sedangkan untuk *subjectivity* sendiri merupakan sebuah fungsi yang digunakan dalam melihat nilai dari sebuah teks [3].

#### 2. Labeling

*Labeling* ini didapatkan dari hasil *polarity* dan *subjectivity*. Semakin tinggi *subjectivity*, maka dapat dikatakan opini. Namun, jika semakin tinggi *polarity* maka dikatakan bahwa itu merupakan emosi positif. Dari nilai *polarity* ini lah dapat dikategorikan kedalam positif, netral, dan negatif [3].

#### 3. Result

Tahap *result* merupakan tahap dimana melakukan penggambaran terhadap hasil yang didapatkan dari tahap *labeling*. Pada tahap ini, akan dihitung jumlah sentimen positif, netral, dan negatif yang didapatkan. Setelah itu, akan dihitung persentasenya untuk setiap sentimen dan divisualisasikan dengan menggunakan *plot*.

#### 4. Evaluasi

Tahap yang terakhir adalah evaluasi dari data. Pada data ini dilakukan *training* dan *testing* data. Proses *training* dilakukan untuk membangun model, sedangkan proses *testing* dilakukan untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses *training* [9]. Pada data ini dilakukan 70% *training* dan 30% *testing*. Kemudian, didapatkan hasil dari keakuratan model yang dipakai. Berikut merupakan acuan untuk kategori keakuratan model.

Tabel 1. Kategori Keakuratan Model

Nilai Akurasi	Kategori
90% - 100%	Sangat Baik
80% - 90%	Baik
70% - 80%	Cukup
60% - 70%	Kurang
50% - 60%	Gagal

Akurasi merupakan sebuah pengukuran yang banyak dilakukan dalam pengujian model algoritma *machine learning*. Kemudian, yang terakhir adalah menentukan akurasi model dengan menggunakan

metode *confusion matrix*. Metode ini digunakan dalam melakukan pengukuran dan mengevaluasi kinerja model. Berikut merupakan acuan untuk mengkonfigurasi *confusion matrix*.

Tabel 2. *Confusion Matrix*

Prediksi	Sebenarnya		
	Positif	Netral	Negatif
Positif	TP	FP	FP
Netral	-	TP	-
Negatif	FP	FP	TP

Untuk mengukur ketepatan klasifikasi, jumlah pada setiap kelas prediksi dan aktual perlu untuk diketahui. Hal tersebut dapat dilihat sebagai berikut.

- *True Positive* (TP): jumlah ulasan dengan sentimen positif yang tepat terprediksi dalam kelas positif.
- *True Negative* (TN): jumlah ulasan dengan sentimen negatif tepat terprediksi dalam kelas negatif.
- *False Positive* (FP): jumlah ulasan dengan sentimen negatif yang terprediksi dalam kelas positif.
- *False Negative* (FN): jumlah ulasan dengan sentimen positif yang terprediksi dalam kelas negatif.

Setelah itu, hasil tersebut akan digambarkan dalam matriks *heatmap* untuk memudahkan dalam melakukan analisis hasil akhir. Dalam rumus matematis, pengukuran akurasi model ini dapat dituliskan dalam persamaan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN}$$

#### 4. HASIL DAN PEMBAHASAN

Berdasarkan metodologi yang dilakukan, berikut merupakan hasil dan pembahasan dari data yang diolah.

##### Pengumpulan Data Komentar

Pengumpulan data dilakukan dengan *crawling* yang menggunakan tools *tweet-harvest*. Teknik ini memerlukan beberapa variabel tambahan seperti *twitter token* dan *node.js* serta *tweet-harvest* versi terbaru untuk mengekstrak komentar di X. Adapun komentar yang diambil berasal dari enam unggahan di X, yang berkaitan tentang pembahasan mengenai *DeepFake*, yang masing-masing unggahan tersebut terdiri dari 30-150 komentar. Data yang terkumpul kemudian digabungkan menjadi satu berkas *csv*, seperti yang dapat dilihat di Gambar 2, untuk memudahkan analisis yang akan dilakukan.

```

0      @detikcom klo kami warga indonesia paham sih p...
1      @detikcom Di tik tod bngak pendukung si tri be...
2      @detikcom tp sebenarnya yaa andai pun ini buka...
3      @detikcom pidato bahasa inggris aja masih pela...
4      @detikcom Kasian pak de
      ...
307    @republikaonline Peliharaan penjahat mau bikin...
308    @republikaonline Orang awam tetap kena nih klw...
309    @republikaonline Menguji nyali pak Kapolri @Li...
310    @republikaonline @didik_bung Gerombolan cebong...
311    @republikaonline Sundul tagar #TangkapAbuJanda...
Name: full_text, Length: 312, dtype: object

```

Gambar 2. Hasil dari *Data Crawling*

##### Preprocessing Data

Selanjutnya, sebanyak 312 komentar yang telah didapatkan akan diolah dengan pra-pemrosesan data menggunakan teknik *Case Folding*, *Cleaning*, *Tokenizing*, *Stemming*, dan *Filtering (Stopword Removal)* secara berurutan. Hal ini dilakukan untuk memudahkan analisis pembagian sentimen yang akan dilakukan. Untuk hasil pra-pemrosesan data dapat dilihat pada Tabel 3 di bawah.

Tabel 3. Hasil Pra-pemrosesan data

Langkah	Contoh Sebelum	Contoh Sesudah
<i>Case Folding</i>	klo kami warga indonesia paham sih	klo kami warga indonesia paham sih
	Peliharaan penjahat mau bikin kejahatan	peliharaan penjahat mau bikin kejahatan
	Kasian pak de	kasian pak de
	Orang awam tetap kena nih	orang awam tetap kena nihb h
	Menguji nyali pak Kapolri	menguji nyali pak kapolri
<i>Cleaning</i>	312 Data	306 Data
<i>Tokenizing</i>	klo kami warga indonesia paham sih	[klo, kami, warga, indonesia, paham, sih]
	peliharaan penjahat mau bikin kejahatan	[peliharaan, penjahat, mau, bikin, kejahatan]
	kasian pak de	[kasian, pak, de]
	orang awam tetap kena nih	[orang, awam, tetap, kena, nih]
	menguji nyali pak kapolri	[menguji, nyali, pak, kapolri]
<i>Stemming</i>	[klo, kami, warga, indonesia, paham, sih]	[klo, kami, warga, indonesia, paham, sih]
	[peliharaan, penjahat, mau, bikin, kejahatan]	[pelihara, jahat, mau, bikin, jahat]
	[kasian, pak, de]	[kasi, pak, de]
	[orang, awam, tetap, kena, nih]	[orang, awam, tetap, kena, nih]
	[menguji, nyali, pak, kapolri]	[uji, nyali, pak, kapolri]

<i>Filtering</i>	[klo, kami, warga, indonesia, paham, sih]	klo warga indonesia paham sih
	[pelihara, jahat, mau, bikin, jahat]	pelihara jahat bikin jahat
	[kasi, pak, de]	kasi de
	[orang, awam, tetap, kena, nih]	orang awam kena nih
	[uji, nyali, pak, kapolri]	uji nyali kapolri

## SVM

Data yang telah melewati *Preprocessing Data* mengkategorikan label tweet tergolong positif, netral, atau negatif, maka nilai polarity dan subjectivity setiap kata dihitung.

	full_text	Polarity	Subjectivity
0	klo warga indonesia paham sih ga pake alat tel...	0.0000	1.000
1	tik tod bngak dukung si tri muncul	0.0000	1.000
2	tp sebenarnya yaa andai hoax emang kaloo bahas...	-0.2732	0.792
3	pidato bahasa inggris aja pelan2 mandarinada a...	0.0000	1.000
4	kasi de	0.0000	1.000

Gambar 3 Nilai *Polarity* dan *Subjectivity*

Selanjutnya, hasil nilai tersebut akan digunakan untuk memberikan label pada setiap baris.

	full_text	Polarity	Subjectivity	Analysis
0	klo warga indonesia paham sih ga pake alat tel...	0.0000	1.000	Netral
1	tik tod bngak dukung si tri muncul	0.0000	1.000	Netral
2	tp sebenarnya yaa andai hoax emang kaloo bahas...	-0.2732	0.792	Negatif
3	pidato bahasa inggris aja pelan2 mandarinada a...	0.0000	1.000	Netral
4	kasi de	0.0000	1.000	Netral

Gambar 4 Hasil *Labelling*

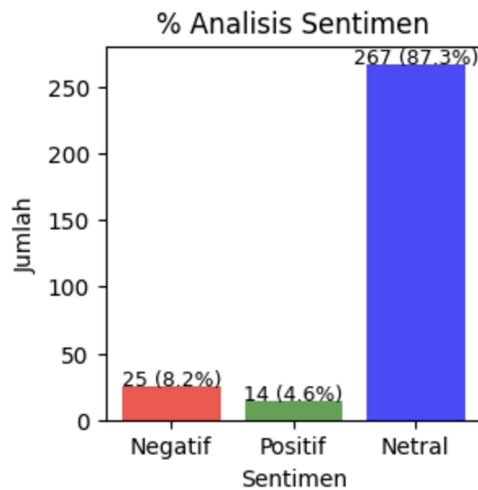
## Analisis Sentimen

Proses Analisis Sentimen dilakukan dengan menggunakan NLP dan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi. Hasil dari tahap SVM akan di *grouping* sesuai kategori 'Analysis' dengan total jumlah pada Gambar 5.

Jumlah Negatif: 25  
Jumlah Positif: 14  
Jumlah Netral: 267  
Jumlah Invalid: 0

Gambar 5 Jumlah per kategori

Berikut ini merupakan plot yang memvisualisasikan total kategori sentimen dalam bentuk diagram bar yang ditunjukkan pada Gambar 6.



Gambar 6 Hasil Klasifikasi

Berdasarkan hasil dari klasifikasi diatas, maka hasil analisis sentimen masyarakat terhadap teknologi deepfake pada aplikasi X dapat dilihat pada Tabel 4.

Tabel 4 Persentase Hasil Analisis Sentimen

Sentimen (%)		
Positif	Netral	Negatif
4,6%	87,3%	8,2%

Hasil yang diperoleh untuk analisis sentimen paling tinggi adalah sentimen netral dengan persentase 87,3%, diikuti dengan sentimen negatif dengan persentase 8,2%, dan sentimen positif dengan persentase 4,6%.

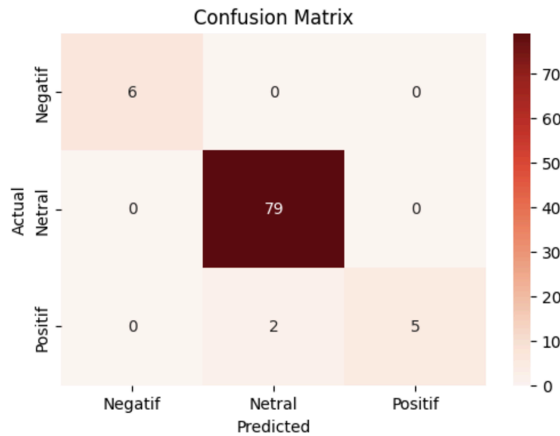
## Evaluasi

Tahap terakhir yaitu dengan menghitung akurasi model NLP dengan evaluasi kinerja menggunakan 70% data *training* dan 30% data *testing*. Berikut merupakan hasil perhitungan Precision, Recall, dan F1-Score beserta heatmap *confusion matrix*.

	precision	recall	f1-score	support
Negatif	1.00	1.00	1.00	6
Netral	0.98	1.00	0.99	79
Positif	1.00	0.71	0.83	7
accuracy			0.98	92
macro avg	0.99	0.90	0.94	92
weighted avg	0.98	0.98	0.98	92

Gambar 7 Hasil Perhitungan Precision, Recall, dan F1-Score





Gambar 8 Heatmap *Confusion Matrix*

Berdasarkan *Confusion Matrix* tersebut, diketahui terdapat:

- **Klasifikasi benar:** Model berhasil memprediksi 6 data negatif, 79 data netral, dan 5 data positif.
- **Klasifikasi salah:** Model gagal memprediksi 2 data positif menjadi data netral.

## 5. KESIMPULAN

Analisis sentimen publik terhadap penggunaan teknologi *deepfake* pada era digital berdasarkan aplikasi X diperoleh data sebanyak 312 kolom komentar menggunakan teknik *crawling*. Proses analisis menghasilkan keluaran berupa 25 komentar negatif, 14 komentar positif, dan 267 komentar netral. Evaluasi kinerja algoritma SVM dengan perbandingan 70% data *training* dan 30% data *testing* menghasilkan akurasi data sebesar 98% dimana akurasi dapat dikategorikan sebagai Sangat Baik sehingga dapat disimpulkan bahwa sentimen masyarakat Indonesia terhadap penggunaan *deep fake* sebagian besar memberikan tanggapan netral dengan jumlah 267 dari 312 komentar. Ketidakakuratan hasil dari penelitian ini disebabkan oleh beberapa faktor seperti perbedaan persepsi ketika memberi label pada setiap komentar serta terbatasnya kosakata pada *repository* Github. Selain itu, penelitian ini hanya menggunakan kurang dari seribu data komentar. Hal ini disebabkan oleh sangat minimnya pembahasan mengenai Deepfake di X oleh pengguna media sosial di Indonesia, yang pada akhirnya juga menjadi salah satu faktor dalam validitas sumber data. Dapat diperhatikan juga terdapat beberapa komentar yang tidak terlalu relevan dengan topik yang sedang dibahas sehingga dapat mempengaruhi hasil analisis yang dilakukan. Lalu, perbedaan konteks pada data dari setiap unggahan di X yang kami ambil juga menjadi salah satu faktor kurangnya kualitas model yang terbentuk. Kedepannya, penelitian ini dapat dilanjutkan dengan sumber data yang valid dan audiens yang lebih luas.

## REFERENSI

- [1] Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal Bumigora Information Technology (BITE)*, 2(1), 40-44.
- [2] Armadianti, W., Lastono, A. S. B., Putra, F. R., Rakhmawati, N. A. (2024). Citations of publication by user Nbqq79EAAAAJ. Retrieved November 17, 2024, from [https://scholar.google.com/citations?view\\_op=view\\_citation&hl=id&user=Nbqq79EAAAAJ&sortby=pubdate&citation\\_for\\_view=Nbqq79EAAAAJ:v1\\_lew4L6wgC](https://scholar.google.com/citations?view_op=view_citation&hl=id&user=Nbqq79EAAAAJ&sortby=pubdate&citation_for_view=Nbqq79EAAAAJ:v1_lew4L6wgC)
- [3] Baita, A., Pristyanto, Y., & Cahyono, N. (2021). Analisis sentimen mengenai vaksin Sinovac menggunakan algoritma support vector machine (SVM) dan k-nearest neighbor (KNN). *Information System Journal (INFOS)*, 4(2), 42-46.
- [4] Catchmark IT. (2023). *Deepfake Technology and How Businesses Can Protect Themselves*.
- [5] Detty, D., & Tim Penulis. (2023). Pengantar metode analisis sentimen. Penerbit Universitas Gunadarma. Retrieved November 17, 2024, from [https://penerbit.gunadarma.ac.id/wp-content/uploads/2023/09/Pengantar-Metode-Analisis-Sentimen\\_Detty-cs\\_Watermark.pdf](https://penerbit.gunadarma.ac.id/wp-content/uploads/2023/09/Pengantar-Metode-Analisis-Sentimen_Detty-cs_Watermark.pdf)

- [6] Fahmi, R. A. J., Nur, W. M., Canawine, D., Kusumajaya, M. N., Fadhlillah, A. F., Rakhmawati, N. A. (2024). *Analisis Sentimen Masyarakat Terhadap Undang-Undang Perlindungan Data Pribadi Pada Aplikasi X Dengan Metode Support Vector Machine*.
- [7] Ismail, M. A. (2023). Apa itu AI (Kecerdasan Buatan)? Pengertian, kelebihan, kekurangan. Retrieved November 17, 2024, from <https://stekom.ac.id/artikel/apa-itu-ai-kecerdasan-buatan-pengertian-kelebihan-kekurangan#pengertian-ai>
- [8] Jain, A. (n.d.). All About Tokenization, Stop Words, Stemming, and Lemmatization in NLP. Retrieved November 26, 2024, from <https://medium.com/@abhishekjainindore24/all-about-tokenization-stop-words-stemming-and-lemmatization-in-nlp-1620ffaf0f87>
- [9] Jayanti, N. K. D. A., Novianti, K. D. P., & Sumalya, I. W. (2017, February 4). Implementasi metode Support Vector Machine pada sistem pengenalan jejak. *Seminar Nasional Teknologi Informasi dan Multimedia 2017*, STMIK AMIKOM Yogyakarta, 1.2-163–1.2-168.
- [10] Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. (2023). *Deepfake, AI Crime, UU PDP, dan KUHP Baru: Tantangan dan Regulasi di Era Digital*.
- [11] MIT Media Lab. (2020). *Deepfakes Explained: Understanding the Technology and its Impacts*. Massachusetts Institute of Technology.
- [12] Nugroho, K. S. (n.d.). Dasar Text Preprocessing dengan Python. Retrieved November 26, 2024, from <https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe>
- [13] P2K Stekom. (n.d.). *X (media sosial)*. Retrieved November 17, 2024, from [https://p2k.stekom.ac.id/ensiklopedia/X\\_\(media\\_sosial\)](https://p2k.stekom.ac.id/ensiklopedia/X_(media_sosial))
- [14] Satria, H. (n.d.). Crawl Data Twitter Menggunakan Tweet Harvest. November 26, 2024, from <https://helmisatria.com/blog/crawl-data-twitter-menggunakan-tweet-harvest/>
- [15] Setyanto, J., & Sasongko, T. B. (2024). Sentiment Analysis of Sirekap Application Users Using the Support Vector Machine Algorithm.
- [16] Thorp, C. (2023). X (media sosial) dan kebangkitan era baru digital. Retrieved November 17, 2024, from <https://www.bbc.com/indonesia/articles/ckkxrk2q7nwo>
- [17] Universitas Airlangga. (2024). Apa itu deepfake? Ketahui cara kerja dan bahayanya. Retrieved November 17, 2024, from <https://unair.ac.id/apa-itu-deepfake-ketahui-cara-kerja-dan-bahayanya/>