KLASIFIKASI PELANGGARAN UU ITE PADA TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES

Yogiswara Dharma Putra[1)] **I Putu Sugi Almantara**[2)] **I Made Bagus Wahyu Mahendra**[3)] **Ida Bagus Paalakaa RNB**[4)]
Program Studi Teknologi Informasi[1)2)3)4)]
Fakultas Teknik, Universitas Udayana, Badung, Bali [1)2)3)4)]
yogiswaradharmaputra@unud.ac.id[1)], sugialmatara@unud.ac.id[2)]
mahendra.2205551002@student.unud.ac.id[3)] , paalakaa.2205551003@student.unud.ac.id[4)]

ABSTRACT

Serious problems arising from prohibited acts or crimes in the use of information technology have become a major concern in several countries. Even though social media should be a comfortable virtual environment for everyone, in reality many cases of legal violations occur in it. The ITE Law exists as an effort to protect internet users, although the process of determining sanctions for violations requires experts and takes quite a long time. Therefore, an approach is needed to automate the classification of violations based on the articles in the ITE Law. This research aims to classify these violations using the Naive Bayes algorithm, with a case study focused on analyzing comments on the social media platform TikTok. In its experiments, this research grouped violations into five main categories: pornography, defamation, hate speech, online terror, and cyberbullying. The results of this experiment show that the classification model with the Naive Bayes algorithm achieved an accuracy level of 0.83 or 83%. Thus, it is hoped that this information retrieval system will make it easier for social media users to identify violations of the ITE Law that occur in comments on TikTok.

***Keywords** : Classification, Violation, UU ITE, TikTok, Naive Bayes Algorithm, Text Mining*

ABSTRAK

Permasalahan serius yang muncul akibat perbuatan dilarang atau kejahatan dalam penggunaan teknologi informasi telah menjadi perhatian utama di beberapa negara. Meskipun media sosial seharusnya menjadi lingkungan virtual yang nyaman bagi semua orang, kenyataannya banyak kasus pelanggaran hukum terjadi di dalamnya. UU ITE hadir sebagai upaya perlindungan bagi pengguna internet, meskipun proses menetapkan sanksi untuk pelanggaran membutuhkan ahli dan memakan waktu yang cukup lama. Oleh karena itu, diperlukan pendekatan untuk mengotomatisasi klasifikasi pelanggaran berdasarkan pasal-pasal dalam UU ITE. Penelitian ini bertujuan untuk mengklasifikasikan pelanggaran tersebut menggunakan algoritma Naive Bayes, dengan studi kasus terfokus pada analisis komentar di platform media sosial TikTok. Dalam eksperimennya, penelitian ini mengelompokkan pelanggaran ke dalam lima kategori utama: pornografi, pencemaran nama baik, ujaran kebencian, teror online, dan cyberbullying. Hasil dari percobaan ini menunjukkan bahwa model klasifikasi dengan algoritma Naive Bayes mencapai tingkat akurasi sebesar 0.83 atau 83%. Dengan demikian, sistem temu kembali informasi ini diharapkan dapat memberikan kemudahan bagi pengguna media sosial dalam mengidentifikasi pelanggaran UU ITE yang terjadi dalam komentar di TikTok.

Kata kunci : Klasifikasi, Pelanggaran, UU ITE, TikTok, Text Mining, Algoritma Naive Bayes

PENDAHULUAN

Perkembangan teknologi dalam era revolusi industri 4.0 telah mengubah banyak aspek kehidupan manusia, mulai dari sektor pertanian, kesehatan, industri, hingga interaksi sosial. Media sosial menjadi salah satu manifestasi kemajuan teknologi ini, memungkinkan interaksi global melalui platform berbasis internet seperti TikTok, Facebook, Twitter, WhatsApp, dan forum diskusi online. Popularitas media sosial ini tidak hanya terbatas pada hiburan dan ekspresi pribadi, tetapi juga memfasilitasi berbagai kegiatan sehari-hari dengan cara yang lebih bebas dan terbuka untuk berbagi ide dan informasi [1]. Media sosial sangat populer di kalangan masyarakat Indonesia, di mana mereka dapat dengan bebas menyuarakan pendapat dan menyebarkan informasi melalui platform tersebut. Salah satu contoh platform yang digemari di Indonesia adalah TikTok. TikTok sebagai media sosial yang memungkinkan pengguna untuk mengekspresikan diri mereka melalui video pendek dengan berbagai konten kreatif. Pengguna TikTok di Indonesia mencapai 126 juta pengguna yang menjadikan Indonesia sebagai urutan kedua pengguna TikTok terbanyak di dunia [2].

Semakin banyak pengguna TikTok di Indonesia, semakin besar peluang terjadinya kasus pelanggaran di media sosial. Kebebasan berkomentar di *platform* TikTok, seperti halnya di *platform* media sosial lainnya, memberikan ruang bagi ekspresi individu yang luas. Namun, jika tidak diawasi dengan baik, kebebasan ini juga bisa dimanfaatkan untuk menyebarkan komentar yang melanggar hukum atau norma sosial, seperti pornografi, pencemaran nama baik, ujaran kebencian, teror *online*, dan *cyberbullying*. Tidak jarang ditemukan kasus pelanggaran di media sosial khususnya TikTok dengan berkedok bercanda. Terkadang, apa yang dimaksud sebagai bercanda oleh pengguna dapat dianggap sebagai pelanggaran oleh pihak lain, terutama jika komentar yang diberikan sensitif atau menyinggung. Sebagai respons terhadap kejahatan dan pelanggaran yang terjadi, pemerintah Republik Indonesia telah mengeluarkan Undang-Undang No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik, yang dicatat dalam Lembaran Negara Republik Indonesia No. 58 Tahun 2008.

Undang-Undang Republik Indonesia Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE) mengatur tentang berbagai jenis data elektronik seperti teks, suara, dan gambar yang memiliki makna atau dapat dipahami oleh individu yang mampu memahaminya. UU ITE pertama kali disahkan pada tahun 2008 dan mengalami revisi melalui Undang-Undang Nomor 19 Tahun 2016.Termasuk di dalamnya adalah format seperti *electronic data interchange* (EDI), surat elektronik (*email*), telegram, teleks, *telecopy*, serta kombinasi huruf, angka, dan simbol lainnya yang sudah diolah. Di satu sisi, transaksi elektronik merujuk pada tindakan hukum yang dilakukan melalui komputer, jaringan komputer, atau media elektronik lainnya. Undang-Undang ITE berlaku bagi setiap individu yang melakukan tindakan sebagaimana diatur dalam undang-undang ini, baik di dalam maupun di luar wilayah hukum Indonesia, dengan konsekuensi hukum yang dapat berdampak di Indonesia dan merugikan kepentingan nasional [3].

Dalam menetapkan sanksi atas pelanggaran yang terjadi, pentingnya kehadiran seorang ahli yang memahami secara mendalam Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) sangatlah diperlukan. Proses penentuan pasal yang relevan dengan pelanggaran tersebut juga membutuhkan waktu yang cukup lama karena melibatkan analisis yang cermat terhadap konteks kasus [4]. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengotomatisasi proses klasifikasi sanksi untuk pelanggaran berdasarkan ketentuan UU ITE tersebut. Dengan adanya pendekatan otomatisasi ini diharapkan

dapat mempercepat proses penentuan sanksi serta memastikan konsistensi dan akurasi dalam implementasinya.

Berdasarkan konteks di atas, tujuan dari studi ini adalah untuk mengembangkan model klasifikasi yang mampu mengidentifikasi dan mengelompokkan tindakan kriminal serta perilaku yang melanggar UU ITE, terutama di platform media sosial TikTok. Penelitian ini memiliki signifikansi karena jarang ditemukan penelitian yang menciptakan model yang dapat menangani tantangan ini secara efisien. Dalam penelitian ini, penulis akan mengevaluasi kinerja algoritma Naive Bayes karena algoritma tersebut terbukti efektif dalam klasifikasi teks. Diharapkan hasil penelitian ini dapat mendukung penegakan hukum, terutama UU ITE, serta mengurangi insiden kejahatan dan penyalahgunaan teknologi informasi di platform media sosial TikTok.

KAJIAN PUSTAKA

Kajian pustaka adalah kumpulan teori-teori dan referensi yang digunakan sebagai landasan untuk menjawab permasalahan atau menguraikan ide pokok dalam suatu penelitian.

1. Undang-undang ITE

Undang-undang Informasi dan Transaksi Elektronik, atau UU ITE, mengatur mengenai berbagai jenis data elektronik seperti teks, suara, peta, gambar, desain, email, foto, dan lain-lain yang dapat dimengerti oleh penerima informasi [5]. Selain itu, UU ITE juga mengatur mengenai transaksi elektronik, yang merupakan tindakan hukum yang dilakukan melalui komputer, jaringan komputer, atau media elektronik lainnya. UU ITE juga mencakup larangan terhadap beberapa tindakan seperti menyebarkan konten dewasa, perjudian online, pencemaran nama baik, ancaman dan pemerasan, ujaran kebencian, terorisme online, peretasan akun media sosial, penyebaran informasi palsu atau hoaks, serta kegiatan lainnya. UU ITE dirancang untuk melindungi keamanan, ketertiban, kesejahteraan, dan kedamaian masyarakat, serta sebagai bentuk perlindungan terhadap ancaman yang bisa merugikan pihak yang dilindungi pemerintah [6].

2. Aplikasi TikTok

TikTok adalah platform jejaring sosial dan video musik yang memfasilitasi pengguna untuk membuat, mengedit, serta membagikan klip video pendek dengan berbagai filter dan musik sebagai latar belakang. Dengan menggunakan platform ini, pengguna dapat dengan cepat menciptakan video singkat yang kreatif dan secara mudah membagikannya kepada teman serta masyarakat global.Saat ini, TikTok telah menjadi salah satu platform media sosial yang paling populer dan banyak digunakan di berbagai negara. Sebagai platform media sosial yang terbuka untuk semua orang, TikTok memungkinkan pengguna untuk membuat dan membagikan video. Namun, ini juga berarti bahwa beberapa konten yang tidak pantas atau tidak sesuai mungkin muncul. Selain itu, TikTok juga memiliki potensi risiko *cyberbullying* [7].

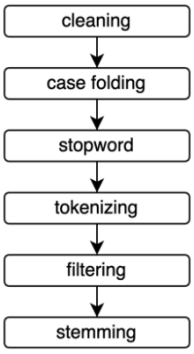
3. Klasifikasi Teks

Klasifikasi teks adalah proses memberikan tag atau kategori pada teks berdasarkan kontennya [8]. Ini dilakukan untuk mengatur, mengelompokkan, dan mengkategorikan informasi yang beragam. Contohnya adalah pengindeksan dokumen berdasarkan kata kunci terkontrol, penyaringan dokumen, pembuatan metadata secara otomatis, dan aplikasi lain yang membutuhkan pengorganisasian dokumen. Klasifikasi teks melibatkan dua tahap utama. Tahap pertama adalah ekstraksi fitur yang penting selama tahap pelatihan, sedangkan tahap kedua melibatkan klasifikasi dokumen setelah melewati pengujian.

4. *Preprocessing*

Preprocessing teks adalah tahap awal dalam pengolahan teks untuk menyiapkan data yang akan diolah lebih lanjut [5]. Langkah ini melibatkan pemisahan sekumpulan karakter berurutan

(teks) menjadi elemen yang lebih bermakna, sambil menghilangkan kata-kata yang tidak relevan untuk membedakan satu dokumen dari dokumen lainnya. Teks *preprocessing* terdapat tahapan yang harus dilalui dapat dilihat sebagai berikut.



Gambar 1. Tahapan Teks *Preprocessing*

Berdasarkan gambar diatas tahapan teks *preprocessing*, akan dilakukan mulai dari tahap *cleaning, case folding, stopword, tokenizing, filtering*, dan *stemming* [9].

a. *Cleaning* adalah sebuah proses menghapus teks atau kata yang tidak relevan untuk mengurangi gangguan dalam proses klasifikasi (ASUS, 2019). Teks atau kata yang dihilangkan merupakan sebuah karakter.

b. *Case folding* adalah sebuah tahapan untuk mengubah huruf dalam dokumen menjadi huruf kecil atau *lowercase*. Selain itu, karakter-karakter non-huruf atau non-angka, seperti tanda baca dan spasi, dianggap sebagai pemisah (*delimiter*).

c. *Stopword* adalah kumpulan teks atau kata yang sering muncul pada sebuah dokumen. Biasanya, *stopword* merujuk pada teks atau kata penghubung sehingga dapat diabaikan dalam proses pengindeksan. Kata-kata ini cenderung kurang deskriptif dan bisa dihapus. Contohnya adalah "dan", "yang", "di", dan sebagainya.

d. *Tokenizing* adalah sebuah proses memisahkan kalimat menjadi sebuah kata dengan menganalisis data, yakni dengan memisahkan setiap kata dan menentukan struktur dari masing-masing kata tersebut.

e. *Filtering* merupakan tahapan untuk mengambil term kunci dari hasil tokenisasi. Hal ini bisa melibatkan penggunaan proses *stoplist* untuk menghapus term yang kurang relevan untuk menyimpan term yang penting. *Stoplist* merujuk kepada term yang kurang deskriptif dan bisa dihapus dari analisis.

f. *Stemming* adalah proses mengurangi variasi kata ke bentuk dasarnya dengan menghapus *suffix* dan *prefix*. Ini membantu dalam pengelompokan kata-kata yang memiliki arti serupa meskipun berbeda dalam bentuk karena adanya imbuhan yang berbeda.

5. Algoritma TF-IDF

Algoritma TF-IDF atau *Term Frequency-Inverse Document Frequency* adalah metode yang memberikan bobot kepada hubungan antara suatu kata (term) dengan dokumen tertentu [10]. Konsep utamanya terdiri dari dua bagian, yaitu *Term Frequency* (TF) yang mengukur seberapa sering suatu kata muncul dalam dokumen. Pendekatan umum untuk menghitung TF adalah dengan membagi jumlah kemunculan kata tersebut dengan jumlah total kata dalam dokumen. Di beberapa situasi, TF bisa disesuaikan dengan menerapkan skema penimbangan yang lebih kompleks. IDF, di sisi lain, mengukur signifikansi suatu kata dalam konteks keseluruhan kumpulan dokumen yang lebih besar. Kata-kata yang jarang muncul di seluruh koleksi dokumen memiliki nilai IDF yang lebih tinggi. IDF dihitung dengan membagi jumlah total dokumen dalam koleksi dengan jumlah dokumen yang mengandung kata tersebut, dan kemudian hasilnya diambil logaritma untuk penyesuaian skala. Berikut adalah rumus algoritma TF-IDF [11].

$$tf = 0{,}5 + 0{,}5 \times \frac{tf}{max(tf)} \quad (1)$$

$$idf_t = log\left(\frac{D}{df_t}\right) \quad (2)$$

$$W_{d.t} = tf_{d.t} \text{ x } idf_{d.t} \quad (3)$$

Dimana,
D = dokumen ke-d
t = term ke-t dari dokumen
W = bobot ke-d terhadap term ke-t
tf = jumlah kemunculan term i pada dokumen
idf = inverse document frequency
df = banyak dokumen yang memiliki term i

6. Metode Naive Bayes Classifier

Klasifikasi Bayes adalah metode statistik untuk mengklasifikasikan data dengan memprediksi probabilitas masuknya data ke dalam kelas tertentu berdasarkan perhitungan probabilitas. Klasifikasi Bayes berdasarkan Teorema Bayes yang ditemukan oleh Thomas Bayes pada abad ke-18. Dalam studi perbandingan algoritma klasifikasi, terdapat metode Bayes yang sederhana yang dikenal sebagai Naive Bayes Classifier. Naive Bayes classifier terbukti memiliki tingkat akurasi dan kecepatan yang tinggi ketika digunakan dalam basis data besar. Metode klasifikasi ini sering digunakan dalam bidang *machine learning* karena kombinasi tingkat akurasi yang tinggi dan perhitungan yang relatif sederhana.

Teorema Bayes merupakan aturan dasar dari algoritma Naive Bayes Classifier yang dapat dilihat pada Persamaan 4 [12].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

Dimana,
$P(H|X)$ = Probabilitas bersyarat H yang diberikan oleh X
$P(X|H)$ = Probabilitas bersyarat X yang diberikan oleh H
$P(H)$= Probabilitas kejadian H
$P(X)$ = Probabilitas kejadian X

Dengan menggunakan Teorema Bayes, penelitian ini akan menerapkan aturan Bayes pada sebuah studi kasus spesifik. Sehingga, Teorema Bayes dapat dirumuskan sebagai berikut.

$$P(C_j|X) = \frac{P(X|C_j)P(C_j)}{P(X)} \quad (5)$$

Dimana C_j merupakan kategori teks atau term yang akan diklasifikasi dan $P(C_j)$ merupakan probabilitas pada kategori teks atau term C_j.

Saat melakukan klasifikasi dokumen teks, pendekatan Bayes akan menentukan kategori dengan probabilitas tertinggi, yaitu C_{MAP} (*Maximum A Posteriori Probability*), menggunakan persamaan berikut.

$$C_{MAP} = argmax \frac{P(X|C_j)P(C_j)}{P(X)} \quad (6)$$

Karena nilai P(X) konstan untuk semua cj, maka persamaan tersebut dapat disederhanakan menjadi.

$$C_{MAP} = argmax\, P(X|C_j)P(C_j) \quad (7)$$

Probabilitas $P(C_j)$ diestimasikan dengan menghitung total dokumen latih pada setiap kategori C_j. Namun, menghitung distribusi $P(X|C_j)$ menjadi sangat sulit dikarenakan jumlah kata atau teks dapat sangat besar. Hal ini menyebabkan jumlah kombinasi posisi kata atau teks yang setara dengan jumlah dari semua kelas atau kategori yang akan diklasifikasikan. Dengan menggunakan algoritma Naive Bayes, mengasumsikan bahwa setiap kata atau teks dalam setiap kategori atau kelas merupakan independen satu dengan yang lainnya, perhitungannya akan disederhanakan dan dinyatakan seperti Persamaan 8.

$$P(X|C_j) = \prod_{i=1}^{n} P(W_i|C_j) \quad (8)$$

Dengan memanfaatkan Persamaan 5, maka Persamaan 8 dijabarkan sebagai berikut.

$$C_{MAP} = argmax \prod_{i=1}^{n} P(W_i|C_j)\, P(C_j) \quad (9)$$

Pada nilai $P(W_i|C_j)$ dan $P(C_j)$ dinilai selama proses latihan dengan menggunakan persamaan sebagai berikut.

$$P(C_j) = \frac{|docs\ j|}{|contoh|} \quad (10)$$

$$P(W_i|C_j) = \frac{1 + n_i}{|C| + n_{(kosakata)}} \quad (11)$$

Dimana,

$P(W_i|C_j)$ = Probabilitas kata W_i pada kategori C_j

$|docs\ j|$ = Probabilitas dokumen pada kategori j

$|contoh|$ = Jumlah seluruh dokumen sampel yang digunakan dalam proses training

n_i = Frekuensi kemunculan kata W_i pada kategori C_j

$|C|$ = Jumlah semua kata pada kategori C_j

$n_{(kosakata)}$ = Jumlah kata yang unik pada semua data training

7. Performance Measure

Langkah terakhir dalam klasifikasi teks adalah pengukuran kinerja atau *performance measure*, di mana evaluasi dilakukan terhadap hasil dari data latih yang akan membandingkan dan menganalisis dari kinerja klasifikasi sebuah teks.

Terdapat metode yang digunakan dalam pengukuran yaitu *precision*, *recall*, *error*, akurasi, dan yang lainnya. Penelitian ini, melakukan evaluasi menggunakan *recall*, *precision*, dan f-*measure*. Hasil dari klasifikasi dokumen untuk setiap kategori ditunjukkan pada Tabel 1 [12].

Tabel 1. Matriks Kebingungan

		Keadaan Data Sebenarnya	
		TRUE	**FALSE**
Hasil Prediksi	**TRUE**	TP (*True Positive*) disebut juga *correct result*	FP (*False Positive*) disebut juga *unexpected result/ false alarm*
	FALSE	FN (*False Negative*) disebut juga *missing result*	TN (*True Negative*) disebut juga *correct rejection*

Pada Tabel 1 menunjukkan hasil klasifikasi dokumen dengan TP (*True Positive*) untuk klasifikasi yang benar dan FP (*False Positive*) untuk yang salah. Dokumen yang tidak terklasifikasi dalam kategori tersebut dibagi menjadi TN (*True Negative*) jika memang bukan anggota kategori, atau FN (*False Negative*) jika seharusnya termasuk dalam kategori tersebut (A, 2023). Keempat parameter ini digunakan untuk menghitung tiga metode evaluasi berikut.

Recall adalah rasio antara jumlah dokumen relevan yang teridentifikasi oleh sistem dibandingkan dengan total jumlah dokumen yang sebenarnya relevan. Rumus *Recall* adalah sebagai berikut.

$$Recall = \frac{TP}{TP + P} \quad (9)$$

Precision adalah sebuah perbandingan antara total jumlah dokumen yang teridentifikasi relevan oleh sistem dengan total dokumen terindentifikasi Precision memiliki rumus yang ditunjukkan pada Persamaan 10.

$$\text{Precision} = \frac{TP}{TP + FN} \quad (10)$$

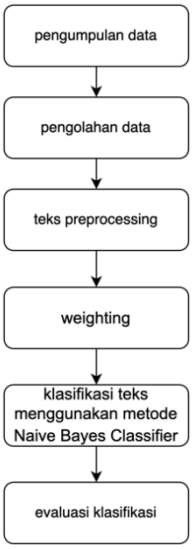
F-measure adalah sebuah nilai yang mencerminkan kinerja dari keseluruhan sistem dan sebuah kombinasi nilai *recall* dan *presis* (Handayani & Pribadi, 2015). Rumus f-measure ditunjukkan pada Persamaan 11.

$$F - measure = \frac{2PR}{P + R} \quad (11)$$

Persentase $recall$, precision, dan $f - measure$ digunakan untuk mengukur kinerja sistem klasifikasi teks otomatis. Semakin tinggi persentase ketiga nilai tersebut, maka akan semakin baik kinerja sistemnya [12].

METODOLOGI PENELITIAN

Penelitian yang dibuat terdiri dari langkah-langkah yang harus dilalui, yaitu pengumpulan dataset, pengolahan dataset, teks *preprocessing* dan klasifikasi terhadap teks menggunakan metode Naive Bayes kemudian tahap terakhir melakukan evaluasi terhadap klasifikasi yang dilakukan. Gambar 2 menunjukkan metodologi penelitian yang akan dilakukan sebagai berikut.



Gambar 2. Metodologi Penelitian

1. Tahap Pengumpulan Data dan Pengolahan Data

Penelitian ini menggunakan komentar teks dari pengguna media sosial TikTok. Data yang digunakan dalam penelitian ini diperoleh melalui proses pengumpulan data yang diperoleh langsung dari komentar di platform TikTok.Data yang dikumpulkan adalah data yang sesuai dan relevan dengan kategori pelanggaran UU ITE yang akan dianalisis. Data yang dikumpulkan tersebut kemudian disatukan dan disaring untuk menghasilkan dataset final yang akan digunakan dalam analisis dan klasifikasi pelanggaran.

Setelah data dikumpulkan, data tersebut diproses untuk mempersiapkan kumpulan data untuk analisis dan klasifikasi pelanggaran. Langkah pertama dalam pengolahan data adalah pembersihan data, di mana data yang dikumpulkan akan diperiksa untuk mengidentifikasi dan menghapus data yang tidak relevan, duplikat, atau tidak berkualitas. Proses ini melibatkan pengecekan manual dan menggunakan algoritma pembersihan data untuk mengidentifikasi dan menghapus entri yang tidak valid atau tidak berguna. Setelah data dibersihkan, langkah selanjutnya adalah penggabungan atau penyatuan data jika diperlukan. Jika data diperoleh dari sumber yang berbeda atau dalam format yang berbeda, data tersebut perlu disatukan menjadi satu dataset tunggal untuk analisis yang konsisten dan lengkap.

2. Tahap *Preprocessing*

Sebelum memasuki tahap klasifikasi, terlebih dahulu dilakukan text preprocessing untuk mereduksi teks dengan menghilangkan kata atau istilah yang tidak memberikan kontribusi signifikan atau mempunyai bobot pada tahap berikutnya. Preprocessing teks ini meliputi pembersihan, pelipatan huruf besar/kecil, *stopwords*, tokenisasi, normalisasi ejaan, pemfilteran, dan *stemming*, yang akan menghasilkan query untuk digunakan pada tahap berikutnya.

3. Tahap *Weighting*

Pada tahap pembobotan, teks diubah menjadi nilai numerik sehingga dapat dihitung dengan algoritma Naive Bayes. Proses ini berlangsung setelah data latih melewati tahap *preprocessing*. Selanjutnya, data pelatihan akan diberi bobot menggunakan metode *Term Frequency-Inverse Document Frequency*.

4. Tahap Klasifikasi Teks Menggunakan Algoritma Naive Bayes

Tahap pada algoritma Naive Bayes, teks diklasifikasi berdasarkan *query* atau data latih yang telah diperoleh sebelumnya. Penelitian terkait algoritma Naive Bayes melibatkan beberapa tahapan, termasuk perhitungan, *conditional probability*, *prior probability* dan menentukan pemilihan kelas berdasarkan nilai maksimum dari kelas yang dipilih.

5. Tahap Evaluasi

Tahap evaluasi melibatkan perbandingan hasil klasifikasi algoritma Naive Bayes dengan menghitung *recall*, *precision*, dan *f-measure*, yang di mana akurasinya dihitung menggunakan matriks kebingungan (*confusion matrix*).

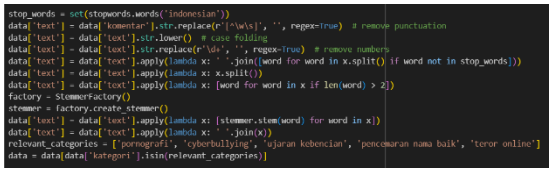
HASIL DAN PEMBAHASAN

Pengklasifikasian kasus kejahatan atau pelanggaran terhadap UU ITE terkait

komentar TikTok, memiliki beberapa tahapan yang akan dilakukan yaitu tahap *preprocessing*, tahap *weighting*, tahap *learning*, dan tahap evaluasi.

1. Preprocessing

Preprocessing merupakan tahapan pertama untuk melakukan pengolahan data sebelum dilakukan pengklasifikasian menggunakan algoritma Naive Bayes. Penelitian menggunakan data komentar sebanyak 120 komentar. Berikut merupakan kode program yang digunakan untuk melakukan *preprocessing* pada dataset.



```
stop_words = set(stopwords.words('indonesian'))
data['text'] = data['komentar'].str.replace(r'[^\w\s]', '', regex=True)  # remove punctuation
data['text'] = data['text'].str.lower()  # case folding
data['text'] = data['text'].str.replace(r'\d+', '', regex=True)  # remove numbers
data['text'] = data['text'].apply(lambda x: ' '.join([word for word in x.split() if word not in stop_words]))
data['text'] = data['text'].apply(lambda x: x.split())
data['text'] = data['text'].apply(lambda x: [word for word in x if len(word) > 2])
factory = StemmerFactory()
stemmer = factory.create_stemmer()
data['text'] = data['text'].apply(lambda x: [stemmer.stem(word) for word in x])
data['text'] = data['text'].apply(lambda x: ' '.join(x))
relevant_categories = ['pornografi', 'cyberbullying', 'ujaran kebencian', 'pencemaran nama baik', 'teror online']
data = data[data['kategori'].isin(relevant_categories)]
```

Gambar 3. Kode Program Preprocessing Data

Berdasarkan gambar di atas, kode program tersebut bertujuan untuk membersihkan dataset dari berbagai elemen yang tidak diperlukan dan dapat mengganggu dalam melakukan klasifikasi. Proses dimulai dengan melakukan *cleaning*, *case folding*, *stopword*, tokeninasi, *filtering*, dan *stemming*. Berikut adalah sampel *preprocessing* pada data komentar.

Tabel 2. *Preprocessing*

No	Sebelum	Sesudah
1	Cewek cantik montok lagi live hot sampai banyak yang like #cantikk	cewek cantik montok live hot sampai banyak like
2	Sumbing Banyak Gaya Najiss	sumbing banyak gaya najis
3	Gua gak boleh ke Bali sama guru gua, disana tempat maksiat nya parah	gua bali guru gua sana tempat maksiat parah
4	Si Spek ani-ani, makanya sering keluyuran malam-malam. Jelas banget dari penampilannya.	spek ani-ani makanya sering keluyur malam jelas banget tampil
5	Izin bapak, nembak kepala anis hukumnya berapa lama ya?	izin bapak tembak kepala anis hukum berapa lama
6	Rata banget dada nya nengg	rata banget dada neng
7	udah pke scincre itu ini tapi kok tetap hitam sih mba	udah pke scincre tapi kok tetap hitam mba
8	Benih-benih teroris... Btw cantiknya dimana coba? Trus tujuannya pake cadar apa?	benih benih teroris cantik coba terus tujuan pake cadar
9	Curiga money laundry. Nggak mungkin bisa beli mobil mewah kayak gitu.	curiga money laundry mungkin beli mobil mewah gitu
10	Aku bakal bikin kamu menyesal seumur hidup!	aku bakal bikin kamu menyesal seumur hidup

2. *Weighting*

Pada tahap pembobotan atau *weighting*, dilakukan perubahan dari teks dibuah menjadi numerik agar dapat dihitung dengan menggunakan algoritma Naive Bayes. Pembobotan dilakukan setelah data komentar selesai melewati tahap *preprocessing*. Selanjutnya data komentar dibobotkan menggunakan metode TF-IDF. Tabel 3 merupakan data *training* yang digunakan dalam proses melakukan pembobotan menggunakan metode TF-IDF.

Tabel 3. Data *Training*

No	Sebelum	Kategori
1	cewek cantik montok live hot sampai banyak like	pornografi
2	sumbing banyak gaya najis	cyberbullying
3	benih benih teroris cantik coba terus	ujaran kebencian

	tujuan pake cadar	
4	curiga money laundry mungkin beli mobil mewah gitu	pencemaran nama baik
5	izin bapak tembak kepala anis hukum berapa lama	terror online

a. Langkah awal dalam pemberian bobot dimulai dengan menghitung *Term-Frequency* (TF), yaitu jumlah kemunculan kata dalam dokumen data latih yang digunakan untuk perhitungan nilai IDF.
b. Langkah kedua proses pemberian bobot dengan menghitung nilai *Inverse Document Frequency*.
c. Langkah terakhir adalah proses pemberian bobot dengan mengalikan nilai TF dengan nilai IDF untuk mendapatkan nilai TF-IDF. Berikut merupakan kode program yang digunakan untuk melakukan pembobotan dengan metode TF-IDF.

```
indo_stopwords = stopwords.words('indonesian')

vectorizer = TfidfVectorizer(stop_words=indo_stopwords, max_features=10000)
X_train_vect = vectorizer.fit_transform(X_train)
X_test_vect = vectorizer.transform(X_test)

feature_names = vectorizer.get_feature_names_out()
```

Gambar 4. Kode Program Pembobotan TF-IDF

Berdasarkan gambar di atas, kode program tersebut digunakan untuk melakukan pembobotan terhadap setiap kata yang terdapat pada dataset komentar. Metode TF-IDF memberikan bobot lebih tinggi pada kata yang sering muncul dalam dokumen tertentu, namun jarang muncul dalam dokumen lainnya. Hal tersebut membantu dalam membedakan antara berbagai teks dalam dataset. Nilai akhir TF-IDF pada dataset dapat dilihat pada Tabel 4 sebagai berikut.

Tabel 4. Nilai TF-IDF Dokumen 1

No	Kata	TF-IDF
1	cewek	0.510258671
2	cantik	0.510258671
3	montok	0.510258671
4	*live*	0.0
5	*hot*	0.467876333
6	sampai	0.0
7	banyak	0.0
8	*like*	0.0

Tabel di atas menunjukkan hasil akhir dari perhitungan nilai TF-IDF untuk *term* dalam Dokumen 1. Dalam tabel tersebut, kata-kata seperti "cewek", "cantik", dan "montok" memiliki nilai TF-IDF sebesar 0.510258671, menunjukkan bahwa kata-kata ini cukup sering muncul dalam Dokumen 1 dan tidak terlalu umum di seluruh koleksi dokumen, sehingga penting dalam konteks Dokumen 1. Kata "*hot*" memiliki nilai TF-IDF 0.467876333, sedikit lebih rendah tetapi tetap signifikan. Sebaliknya, kata-kata seperti "*live*", "sampai", "banyak", dan "*like*" memiliki nilai TF-IDF 0, menunjukkan bahwa meskipun kata-kata ini muncul dalam Dokumen 1, mereka sangat umum di seluruh koleksi dokumen dan tidak memberikan informasi signifikan dalam membedakan Dokumen 1 dari dokumen lainnya.

3. Klasifikasi

Pada tahap klasifikasi, klasifikasi dihitung dengan menggunakan algoritma Naive Bayes. Data *training* yang telah melewati pra-pemrosesan atau preprocessing dan pembobotan yang akan digunakan sebagai data yang akan diklasifikasi pada tahap *testing* dalam menentukan kategori yang sesuai, seperti data *training*, data *testing* juga telah diproses melalui tahap *preprocessing* (Farhan, 2023). Selain itu, proses optimasi model algoritma Naïve Bayes dilakukan dengan menggunakan teknik pencarian grid (*Grid Search*) untuk menemukan parameter terbaik yang meningkatkan performa model. Berikut merupakan kode program dalam membangun model klasifikasi.

```
classifier = MultinomialNB()
param_grid = {'alpha': [0.1, 0.5, 1.0]}
grid_search = GridSearchCV(classifier, param_grid, cv=5, scoring='accuracy')
grid_search.fit(X_train_vect, y_train)

best_classifier = grid_search.best_estimator_
```

Gambar 5. Kode Program Model Klasifikasi Naive Bayes

Kode program di atas bertujuan untuk membangun model klasifikasi menggunakan *Multinomial* Naive Bayes dan teknik *Grid Seacrh* untuk memaksimalkan klasifikasi data komentar yang dilakukan. Berdasarkan perhitungan dataset komentar asli menggunakan algoritma Naive Bayes di Google Colab Notebooks pada klasifikasi pelanggaran UU ITE, diperoleh nilai akurasi sebesar 0.83 atau 83% yang dapat terlihat pada Gambar 6 sebagai berikut.

	precision	recall	f1-score	support
pornografi	1.00	0.80	0.89	5
cyberbullying	0.80	1.00	0.89	4
ujaran kebencian	0.62	1.00	0.77	5
pencemaran nama baik	1.00	0.40	0.57	5
teror online	1.00	1.00	1.00	5
accuracy			0.83	24
macro avg	0.89	0.84	0.82	24
weighted avg	0.89	0.83	0.82	24

Gambar 6. Hasil Nilai Akurasi

Berdasarkan Gambar 6, dijelaskan nilai *precision* yang didapat dari seluruh data *testing* berada dalam rentang nilai 0.62 hingga 1.0. Kemudian nilai *recall* berada dalam rentang nilai 0.40 hingga 1.0. Terakhir pada nilai *f1-score* berada dalam rentang nilai 0.57 hingga 1.0. Dari sini, diperoleh hasil nilai akurasi sebesar 0.83 atau 83%.

4. Evaluasi

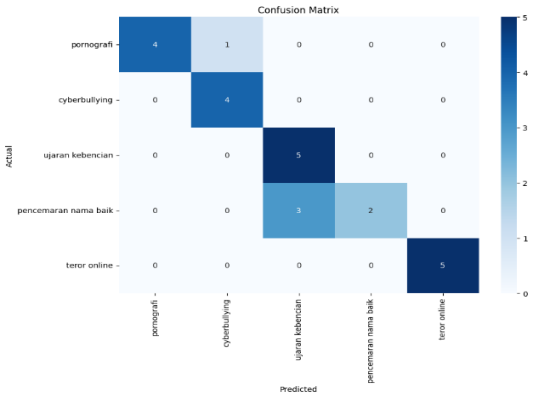
Teknik yang digunakan untuk mengevaluasi klasifikasi dalam penelitian ini adalah dengan menghitung *recall*, *precision*, dan *f-measure*. Teknik ini menggunakan *confusion matrix* sebagai dasar perhitungannya. Berikut merupakan kode program yang digunakan untuk melakukan evaluasi model klasifikasi dengan menampilkan *precision, recall, f-1 score,* dan *accuracy* serta memvisualisasikan hasilnya menggunakan matriks kebingungan atau *confusion matrix.*

```
y_pred = best_classifier.predict(X_test_vect)
print(classification_report(y_test, y_pred, target_names=category_mapping.keys()))

# Visualize results
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=category_mapping.keys(),
            yticklabels=category_mapping.keys())
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

Gambar 7. Kode Program Evaluasi Model

Berdasarkan gambar di atas, kode program tersebut digunakan untuk menampilkan klasifikasi *report* dan menghitung *confusion matrix* yang menghasilkan jumlah prediksi benar dan salah untuk setiap kelas atau kategori yang di tampilkan pada Gambar 8 sebagai berikut.



Gambar 8. *Confusion Matrix*

Secara keseluruhan, *confusion matrix* menunjukkan bahwa algoritma Naive Bayes memiliki performa yang cukup baik dalam mengklasifikasikan komentar, terutama dalam kategori *cyberbullying*, ujaran kebencian, dan teror *online* yang tidak memiliki kesalahan klasifikasi. Namun, terdapat beberapa kesalahan dalam mengklasifikasikan komentar pornografi sebagai *cyberbullying* dan komentar pencemaran nama baik sebagai ujaran kebencian. Hasil ini mengindikasikan bahwa ada ruang untuk perbaikan dalam meminimalkan kesalahan klasifikasi, khususnya untuk kategori pornografi dan pencemaran nama baik.

SIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan sebuah pelanggaran UU ITE pada *platform* TikTok menggunakan algoritma Naive Bayes. Kesimpulan dari penelitian ini menunjukkan bahwa algoritma Naive Bayes, yang menerapkan konsep temu kembali informasi melalui pengolahan data berupa *Text Mining*, efektif dalam mengklasifikasikan komentar di TikTok. Klasifikasi dilakukan ke dalam lima kategori pelanggaran UU ITE dengan

akurasi sebesar 83%, dengan dataset komentar pelanggaran UU ITE yang terdiri dari 120 dataset, yang dibagi menjadi 2 data yaitu 96 data latih dan 24 data uji. Model dibangun menggunakan bahasa pemrograman Python dan Google Colaboratory Notebooks. Performa yang baik dari algoritma Naive Bayes dalam *Machine Learning* menunjukkan hasil akurasi yang tinggi, yang dapat bermanfaat bagi masyarakat dalam mengklasifikasikan pelanggaran pasal UU ITE.

Penelitian ini juga membuka peluang untuk pengembangan lebih lanjut serta penerapan yang lebih luas dalam upaya penegakan hukum di era digital. Adapun saran peneliti, yaitu meningkatkan ukuran dataset dengan lebih banyak data komentar agar model dapat mempelajari pola yang lebih kompleks dan menggeneralisasi dengan lebih baik. Selain itu, diperlukan penanganan ketidakseimbangan data dengan teknik *oversampling* atau *undersampling*, serta penerapan metode seperti SMOTE (*Synthetic Minority Over-sampling Technique*). Selain itu, melakukan optimisasi hyperparameter untuk model yang digunakan dapat memastikan parameter berfungsi secara optimal.

Dengan demikian, sistem temu kembali informasi ini dapat membantu masyarakat atau pengguna media sosial untuk dengan mudah menemukan pelanggaran UU ITE dalam lima kategori pada komentar-komentar di aplikasi TikTok. Implementasi yang lebih luas dari sistem ini tidak hanya berpotensi meningkatkan kesadaran dan pemahaman publik terhadap pelanggaran UU ITE, tetapi juga mendukung penegakan hukum secara lebih efisien dan efektif di era digital. Penelitian ini membuka peluang bagi pengembangan teknologi yang dapat dimanfaatkan oleh penegak hukum, platform media sosial, dan masyarakat umum untuk menciptakan lingkungan digital yang lebih aman dan bertanggung jawab.

REFERENSI

[1] B. Febriyanto, E. Y. Winantika and S. N. Utari, "Peran Media Sosial Dalam Pembentukan Karakter Siswa Di Era DigitaL," *Jurnal Lensa Pendas*, vol. 7, no. 1, pp. 1-14, 2022.

[2] M. S. Pardianti and V. V. S, "Pengelolaan Konten Tiktok Sebagai Media Informasi," *Jurnal Ilmiah Ilmu Komunikasi - Ikon*, vol. 27, no. 2, 2022.

[3] R. Safitri, "Undang-Undang Informasi dan Transaksi Elektronik Bagi Perguruan Tinggi," *Jurnal Sosial & Budaya Syar-i*, vol. 5, no. 3, pp. 198-218, 2018.

[4] A. P. Hesaputra, "Klasifikasi Pelanggaran Undang-Undang ITE pada Twitter Menggunakan LSTM dan BiLSTM," in *Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia*, 2023.

[5] Farhan, Triase and A. M. Harahap, "Penggunaan Algoritma Naive Bayes Dalam Text Mining Untuk Klasifikasi Pasal UU ITE," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 2, pp. 314-322, 2023.

[6] S. R. Ramadoni, R. P. Gegana and K. Sanata, "Sejarah Undang-Undang ITE: Periodisasi Regulasi Peran Negara dalam Ruang Digital," *Jurnal Ilmu Sosial dan Humaniora*, vol. 3, no. 2, pp. 41-58, 2023.

[7] P. N. Rahmana, D. A. P. N and R. Damariswara, "Pemanfaatan Aplikasi Tik Tok Sebagai Media Edukasi di Era Generasi Z," *Jurnal Teknologi Pendidikan*, vol. 11, no. 2, pp. 401-410, 2022.

[8] A. Priyambodo and Prihati, "Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes," *Jurnal Ilmiah Elektronika Dan Komputer*, vol. 13, no. 1, pp. 159-175, 2020.

[9] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine

Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *Jurnal Teknik Informatika - JTI*, vol. 11, no. 2, 2018.

[10] F. Silalahi, A. Hadi and d. S. Widiastuti, "Analisis dan Implementasi Term Frequency - Inverse Document Frequency (TF-IDF) Untuk Filter Etika Buruk Pada Diskusi Online," in *Seminar Nasional Ilmu Komputer*.

[11] D. Septiana and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks," *Jurnal Sistem dan Teknologi Informasi Indonesia - SINTESA*, vol. 1, no. 2, pp. 81-88, 2022.

[12] F. Handayani and F. S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110," *Jurnal Teknik Elektro*, vol. 7, no. 1, pp. 19-24, 2015.