

ANALISIS SENTIMEN TERHADAP KEPERCAYAAN MASYARAKAT PADA PROSES PENEGAK HUKUM DI INDONESIA DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES

Fitri Khoirunnisa^{1*}, Salman Topiq²

^{1,2}Universitas Adhirajasa Reswara Sanjaya, Jl. Terusan Sekolah No. 1-2; (022) 7100124

Received: 28 Juni 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Law;

Sentiment Analysis;

Naïve Bayes;

Correspondent Email:

fitrikhoirunnisa66@gmail.com

Abstrak. Indonesia diakui sebagai negara hukum yang menegakkan kebenaran dan keadilan melalui supremasi hukum. Namun, situasi saat ini kacau dan tidak stabil, termasuk krisis di berbagai sektor termasuk di bidang hukum. Pelaku kejahatan besar sangat sulit ditindak. Tingkat kepercayaan masyarakat terhadap penegakan hukum menurun dari hari ke hari. Saat ini, topik yang menjadi perbincangan hangat di media sosial, terutama di *Twitter*, adalah mengenai penegakan hukum di Indonesia. Banyak pengguna memberikan respon yang positif atau negatif mengenai penegakan hukum di Indonesia. Sehingga, untuk mengetahui pandangan masyarakat terhadap hal tersebut, penelitian ini melakukan analisis sentimen terhadap kepercayaan masyarakat pada proses penegakan hukum di Indonesia dengan menggunakan metode *Naïve Bayes*. Metode *Naïve Bayes* dipilih sebagai metode yang diusulkan dalam penelitian ini karena telah dibuktikan dalam beberapa penelitian terdahulu bahwa metode ini memberikan performansi yang baik terhadap analisis sentimen masyarakat, maka diterapkanlah metode *Naïve Bayes* sebagai implementasi algoritma dalam analisis sentimen, ada beberapa tahapan yang dilakukan dalam proses pengolahan data yaitu tahap data *crawling*, *Pre-processing*, ekstraksi fitur, analisis sentimen dan penerapan algoritma *Naïve Bayes*. Dari hasil yang telah dilakukan maka diperoleh analisis sentimen positif sebesar 41% dan sentimen negatif sebesar 59% serta hasil akurasi dari penerapan algoritma *Naïve Bayes* itu sendiri adalah sebesar 95%.

Abstract. *Indonesia is recognized as a rule-of-law country that upholds truth and justice through the supremacy of law. However, the current situation is chaotic and unstable, including crises in various sectors, including the legal sector. It is very difficult to bring large-scale criminals to justice. The level of public trust in law enforcement is decreasing day by day. Currently, the hot topic of discussion on social media, especially Twitter, is law enforcement in Indonesia. Many users provide positive or negative responses regarding law enforcement in Indonesia. Therefore, to understand public opinion on this matter, this research conducts sentiment analysis on public trust in the law enforcement process in Indonesia using the Naïve Bayes method. The Naïve Bayes method is chosen as the proposed method in this study because it has been proven in several previous studies to provide good performance for analyzing public sentiment, so the Naïve Bayes method is applied as an algorithm implementation in sentiment analysis, and several stages are carried out in the data processing process, namely data crawling, pre-processing, feature extraction, sentiment analysis, and Naïve Bayes algorithm implementation. From the results obtained, a positive sentiment analysis of*

41% and a negative sentiment analysis of 59% were obtained, and the accuracy of the Naïve Bayes algorithm implementation itself is 95%.

1. PENDAHULUAN

Berdasarkan Undang-Undang Dasar Negara Republik Indonesia, Indonesia diakui sebagai negara hukum. Konsep negara hukum tersebut merujuk pada negara yang menegakkan kebenaran dan keadilan melalui supremasi hukum [1]. Saat ini, Indonesia sedang menghadapi situasi yang tidak stabil dan kacau, termasuk krisis yang terjadi di berbagai sektor termasuk di bidang hukum. Meskipun hukum diharapkan dapat memberikan keadilan bagi masyarakat, kenyataannya sebaliknya. Penegakan hukum hanya efektif untuk kasus-kasus kecil yang dilakukan oleh masyarakat kecil. Sementara itu, pelaku-pelaku kejahatan besar seperti korupsi, kolusi, dan nepotisme yang biasa disebut sebagai penjahat berkerah putih sangat sulit untuk ditindak [2].

Menurut Roscoe Pound, seorang tokoh terkemuka dalam bidang *Sociological Jurisprudence*, hukum yang baik adalah hukum yang sejalan dengan kehidupan hukum yang berlaku dalam masyarakat. Dengan kata lain, hukum yang efektif dan adil adalah hukum yang mampu mengakomodasi kebutuhan masyarakat dan mampu menyelesaikan masalah-masalah yang timbul di dalam masyarakat [3]. Pada tingkat kepercayaan masyarakat terhadap penegakan hukum di negara ini menurun dari hari ke hari. Semakin meningkatnya tingkat ketidakpuasan masyarakat terhadap penegakan hukum di negara ini, terlihat jelas dari hasil survei yang dilakukan oleh Lembaga Survei Indonesia. Hasil survei tersebut menunjukkan bahwa tingkat ketidakpercayaan publik terhadap penegakan hukum saat ini mencapai 56 persen dan hanya 29,8 persen yang menyatakan puas. Hal ini perlu dibandingkan dengan survei sebelumnya yang menunjukkan bahwa hanya 22,6 persen yang mengatakan penegakan hukum di era pemerintahan sekarang ini lebih baik dari pemerintahan sebelumnya [4].

Dengan menggunakan *platform* jejaring sosial, kita dapat mengetahui pandangan baik maupun buruk [5] salah satu sarana yang digunakan individu untuk mengekspresikan ide, gagasan, atau opini mereka dan memperlihatkannya kepada orang

lain adalah melalui media sosial. *Twitter* adalah salah satu jenis media sosial yang sering digunakan oleh masyarakat [6]. Banyaknya pengguna masyarakat dalam menggunakan *Twitter* sehingga dapat dimanfaatkan untuk mengetahui sentimen masyarakat mengenai penegakan hukum di Indonesia.

Meskipun penentuan apakah suatu *tweet* mengandung sentimen negatif, positif, atau netral bisa dilakukan secara manual, namun dengan banyaknya pengguna *Twitter* maka opini yang dihasilkan juga semakin banyak. Hal ini dapat memakan waktu dan usaha yang lebih banyak. Oleh karena itu, diperlukan sebuah mesin yang dapat secara otomatis menganalisis *tweet* dan mengklasifikasikannya ke dalam kategori sentimen positif dan negatif. Proses analisis otomatis dibutuhkan karena metode manual sudah tidak efektif lagi mengingat jumlah *tweet* yang semakin banyak [7]. Maka dengan itu dapat diterapkannya teknik *text mining* yang dapat digunakan untuk memenuhi kebutuhan dalam mengelompokkan dokumen dengan berbagai jenis konten [5].

Klasifikasi *Naïve Bayes* (NB) adalah sebuah algoritma pembelajaran mesin yang berdasarkan teori *Bayes*. Algoritma ini menggunakan asumsi yang cukup kuat bahwa pengaruh dari nilai atribut terhadap suatu kelas yang diberikan adalah independen atau tidak bergantung pada kelas atau atribut lainnya. Teori *Bayes* sendiri merupakan sebuah teori statistik klasifikasi yang memungkinkan untuk memprediksi probabilitas atribut kelas anggota berdasarkan data yang telah tersedia [8]. Algoritma *Naïve Bayes* ini dianggap tepat untuk digunakan dalam menganalisis sentimen yang terdapat pada dataset [5].

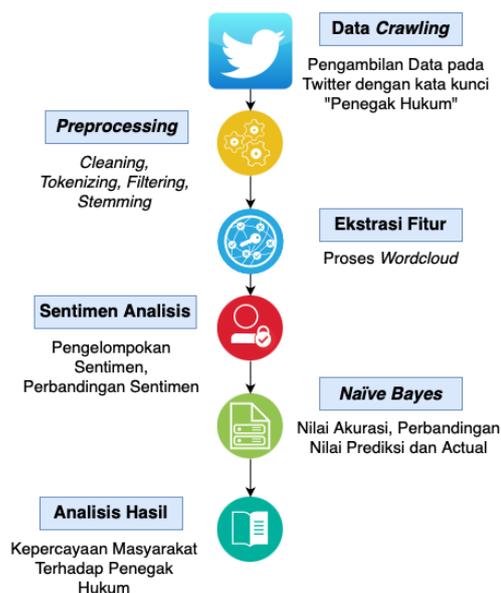
Metode *Naive Bayes* dipilih sebagai metode yang diusulkan dalam penelitian ini karena telah dibuktikan dalam beberapa penelitian terdahulu bahwa metode ini memberikan performansi yang baik terhadap analisis sentiment masyarakat. Beberapa peneliti yang dimaksud antara lain yaitu penelitian yang dilakukan oleh Fajar Ratnawati [9] dengan tema penelitian Implementasi Algoritma *Naïve Bayes* terhadap Analisis

Sentimen Opini Film Pada Twiter dari penelitian yang telah dilakukan di dapatkan hasil akurasi sebesar 90%. Penelitian selanjutnya yang dilakukan oleh Fira Fathonah dan Asti Herliana [5] dengan penelitiannya yaitu Penerapan *Text Mining* Analisis Sentimen Mengenai Vaksin Covid-19 Menggunakan Metode *Naïve Bayes* hasil dari penelitian yang dilakukan yaitu didapatkan analisis sentiment dari social media *Twitter* terkait Vaksin Covid-19 didapatkan persentase *accuracy* sebesar 100%. Selanjutnya terdapat penelitian dari Dedi Darwis dkk [10] dengan artikel yang bertemakan mengenai Penerapan Algoritma *Naïve Bayes* Untuk Analisis Sentimen Review Data *Twitter* BMKG Nasional sehingga didapatkanlah hasil pengujian dengan menggunakan metode *Naïve Bayes* yang dilakuakn yaitu nilai akurasinya untuk klasifikasi yaitu sebesar 69.97%.

Dari paparan penelilitan terdahulu maka dapat dibuktikan bahwa algoritma *Naïve Bayes* tepat dilakukan untuk analisis sentimen masyarakat maka dengan itu penelitian ini bertujuan untuk mencari komentar positif dan negatif pada data *Twitter* dengan *keyword* Penegak Hukum serta algoritma yang digunakan yaitu menggunakan algoritma *Naive Bayes*. Selain itu, penelitian ini juga bertujuan untuk menghitung jumlah data yang termasuk dalam kategori positif dan negatif dengan menggunakan algoritma *Naïve Bayes*, serta menghasilkan perbandingan nilai sentimen. Proses pengolahan data dan analisis sentimen dilakukan menggunakan bahasa R dengan *tools* RStudio.

2. METODE PENELITIAN

Dalam penelitian, metode diperlukan untuk memberikan struktur dan memastikan bahwa hasil penelitian sesuai dengan tujuan yang ditetapkan[11]. Pada Gambar 1 menunjukkan tahapan-tahapan penelitian yang perlu dilakukan dalam penelitian yang akan dilakukan.



Gambar 1. Tahapan Penelitian

2.1. Data Crawling Twitter

Crawling adalah sebuah kegiatan yang dilakukan untuk membuat salinan dari *World Wide Web* dengan mengambil bagian-bagian yang dianggap relevan dan penting. Proses ini biasanya dilakukan oleh program komputer yang disebut sebagai "*web crawler*" atau "*spider*" [11]. *Crawling* sendiri dapat diartikan sebagai suatu yang dilakukan untuk mendapatkan dataset yang terdiri dari *tweet*, yang kemudian akan dijadikan input pada tahap *text preprocessing* [12].

Salah satu jenis media sosial yang menyediakan API untuk memungkinkan pengguna mengakses data mereka adalah *Twitter*. Data yang umumnya diambil dari *Twitter* adalah teks postingan yang dikenal sebagai cuitan atau *tweets* [13]. Pada penelitian ini digunakanlah data ulasan *Twitter* dengan kata kunci yang dicari yaitu penegak hukum.

2.2. Pre-processing

Proses yang akan dilakukan yaitu menggunakan *text mining*, tujuan dari *text mining* adalah menghasilkan informasi dari satu set dokumen dengan memproses, mengelompokkan, dan menganalisis data-data yang tidak terstruktur dalam jumlah besar. Teknik *text mining* memungkinkan untuk mendapatkan informasi yang berguna dari serangkaian dokumen dengan sumber data

pada teks yang memiliki format yang tidak terstruktur [14].

Data yang terkumpul adalah data yang tidak terstruktur sehingga memerlukan pengolahan *Pre-Processing*. Proses ini sangat penting dan tidak dapat diabaikan [15]. Proses *pre-processing* data ini sangat penting untuk menyelesaikan berbagai masalah seperti data yang tidak akurat, data yang redundan, data yang hilang, dan sebagainya [16].

Proses *text preprocessing* digunakan untuk menghilangkan tanda baca, angka, simbol, tautan *URL*, dan nama pengguna dari teks. Pada data yang berasal dari *Twitter*, seringkali terdapat *noise* atau data yang masih kotor karena terdapat karakter yang tidak diperlukan untuk proses klasifikasi, seperti tanda baca, *mention*, *hashtag*, angka, username, alamat *URL*, dan *retweet text*. Oleh karena itu, tahap *text preprocessing* sangat penting untuk membersihkan dan mempersiapkan data sebelum dijadikan input pada tahap selanjutnya [13].

Adapun beberapa tahapan yang dilakukan dalam proses *preprocessing*, yaitu :

1. *Cleaning* :

Proses pembersihan atau *cleaning* dilakukan dengan mengikuti tahapan-tahapan proses data *mining*, sehingga tidak semua tabel yang tersedia di dalam setiap basis data akan digunakan. Dalam proses ini, kata-kata yang tidak diperlukan akan dihilangkan, serta simbol-simbol seperti (), titik, koma, dan sebagainya juga akan dihapus. Hal ini dilakukan untuk mempersiapkan data sebelum diolah lebih lanjut, sehingga dapat menghasilkan hasil yang akurat dan relevan [17].

Pada tahap ini dilakukan juga proses *case folding*, *case folding* digunakan untuk merubah semua huruf dalam kalimat *tweet* menjadi huruf kecil, sehingga tidak terdapat perbedaan antara huruf besar dan kecil. Contohnya, kata "KUNCI" akan diubah menjadi "kunci", dan mencoba memasukkan kata "KunCI" atau "KuNci" akan tetap menghasilkan hasil pencarian yang sama dengan "kunci". Proses ini dilakukan agar lebih mudah dalam mengakses dan memproses data *tweet* dengan hasil yang konsisten dan akurat [13].

2. *Tokenizing*

Tokenizing merupakan suatu proses dalam pengolahan bahasa alami di mana teks atau string input dipotong atau dibagi menjadi potongan-potongan kata yang terpisah. Potongan-potongan kata ini kemudian disebut sebagai token, yang merupakan unit dasar dalam pemrosesan teks. Proses *tokenizing* sangat penting dalam analisis teks dan pemrosesan bahasa alami karena memungkinkan komputer untuk memahami dan memproses teks dengan cara yang lebih terstruktur dan terorganisir. Dalam beberapa kasus, proses *tokenizing* juga dapat melibatkan penghapusan karakter atau tanda baca yang tidak diperlukan, seperti spasi, tanda koma, titik, atau tanda kurung [15].

3. *Filtering*

Filtering adalah sebuah proses penting dalam *text mining* yang dilakukan untuk membuang kata-kata yang tidak memiliki makna atau yang sering disebut sebagai *stopword*. *Stopword* biasanya merupakan kata-kata umum seperti "dan", "atau", "di", "ke", dan sejenisnya yang tidak memberikan kontribusi signifikan dalam analisis teks. Dengan menghilangkan *stopword*, dapat membantu meningkatkan kualitas dan efisiensi dari hasil analisis teks yang dilakukan. Proses *filtering* dilakukan setelah tahap *preprocessing* dalam *text mining*, dan merupakan salah satu langkah penting untuk mempersiapkan data sebelum masuk ke tahap selanjutnya [15].

4. *Stemming*

Stemming adalah proses menghilangkan imbuhan pada kata untuk menemukan akar kata. Hal ini dilakukan untuk mengurangi jumlah kata dalam dokumen atau korpus, mempermudah analisis data, dan meningkatkan efisiensi pemrosesan teks. Meskipun hasil *stemming* tidak selalu tepat, teknik ini masih sering digunakan bersama dengan teknik lain untuk menghasilkan hasil yang lebih akurat [18].

2.3. *Ekstraksi Fitur*

Pada proses ekstraksi fitur dilakukan *word cloud*, *word cloud* dianalisis untuk mengetahui kata-kata yang sering disebutkan dalam konteks penegakan hukum. Setiap sentimen pada *word cloud* dibandingkan untuk

menunjukkan kata-kata yang paling umum digunakan pada *tweet* dengan sentimen positif atau negatif [19].

2.4. Analisis Sentimen

Analisis sentimen adalah teknik yang menggabungkan beberapa bidang seperti *Natural Language Processing* (NLP), *Information Retrieval* (IR), dan *Data Mining* (DM) untuk mengolah atau menganalisis opini, sentimen, dan emosi yang diekspresikan dalam bentuk teks pada suatu entitas [6]. Analisis sentimen juga dapat bertujuan untuk mengolah teks tak teratur menjadi informasi terstruktur berupa ulasan atau opini. Hal ini dilakukan untuk mengubah informasi yang tidak terstruktur menjadi terstruktur, seperti pada format forum, *blog*, media sosial, atau situs yang mengandung ulasan [20].

2.5. Algoritma Naïve Bayes

Algoritma *Naïve Bayes* merupakan sebuah teknik klasifikasi probabilistik yang sangat umum digunakan dalam bidang *Data Mining* dan *Text Mining*. Metode ini didasarkan pada teorema *Bayes*, di mana suatu hipotesis diperbarui berdasarkan bukti baru yang muncul. Algoritma *Naïve Bayes* dikenal sangat efektif dalam mengklasifikasikan data dengan cepat dan akurat, terutama pada kasus di mana jumlah fitur sangat besar dan kompleks. Karena kelebihan tersebut, algoritma ini banyak digunakan dalam berbagai aplikasi, termasuk analisis sentimen, klasifikasi dokumen, deteksi *spam*, dan sebagainya [10].

2.6. Analisis Hasil

Pada hasil akhir yang diterapkan maka akan disusun terhadap keterkaitan kepercayaan masyarakat terhadap penegak hukum yang kemudian diukur dengan analisis sentimen dengan penerapan algoritma *Naïve Bayes*.

3. HASIL DAN PEMBAHASAN

Setelah dilakukan pembahasan pada latarbelakang serta metode yang akan diterapkan, maka pada bagian ini yaitu melakukan implementasi terhadap metode penelitian yang telah dibuat. Dalam proses pengolahan data sentimen analisis ini menggunakan Bahasa R dengan *tools* yang diterapkan yaitu RStudio. Adapun beberapa

proses tahapan yang dilakukan dalam pengolahan data, yaitu :

3.1. Data Crawling

Data yang digunakan dalam penelitian ini berupa ulasan dari *tweet* yang diposting oleh pengguna *Twitter* mengenai Penegakan Hukum. Data tersebut diambil menggunakan fitur #TAGS selama periode 14 hari dan terkumpul sebanyak 4165 data *tweet* mentah dengan dengan 18 atribut, data yang akan dilakukan dalam pengolahan yaitu pada atribut *text*.

Setelah data terkumpul, maka akan dilakukan proses penentuan sentimen berdasarkan makna yang terkandung didalamnya, kemudian dalam prosesnya akan dikategorikan menjadi ulasan positif, atau negatif. Pada Gambar 2 merupakan data *text* mentah yang belum dilakukan pembersihan data.

Tabel 1. *Sample Data Mentah Hasil Crawling Ulasan Twitter*

text
RT @henk_batuwael: @jokerjutex Apakah ini suatu kebanggaan bagi aparat penegak hukum???? Tanya Pak @ListyoSigitP
RT @ChHuda5: Ini apaan sih Jendral @ListyoSigitP @DivHumas_Polri koq ngeri gini Itu mereka bisa bebas pamer senjata dimedia apa gada tindakan dr aparat penegak hukum tah pak @mohmahfudmd Anak bangsa butuh ketenangan & kenyamanan Pak @jokowi Kpn majunya bangsa ini klo yg begini dbiarkan ! https://t.co/IZVBbSLQS2
@__AnakKolong @jokowi @ListyoSigitP Yg berdosa/dilarang itu kejahatan ² perbuatan tdk menyenangkan orang lain, pengancaman, kebencian, permusuhan, intoleran, radikal/terorisme. Giat ibadah agama ² di dunia melakukan kebaikan ² bukan kejahatan. Yg melarang ibadah agama layak dihukum di dunia akhirat oleh penegak hukum.
RT @N1ELSEN: @budimandjatmiko Saya tau kasus berbau sara pasti bakalan rumit buat pak @ListyoSigitP . Cuman pak ini bom waktu pak, ga ditindak sekarang kedepannya

<p>pasti jadinya chaos. Tolong pak jaga WIBAWA POLRI yang masih kami percaya sebagai penegak hukum.</p>
<p>RT @BlueNeo81: . Toleransi itu udh diajarkan di keluarga SD SMP SMA ampe kuliah Klo dgn ada peristiwa itu kita bicarakan lagi Toleransi ITU NAMANYA BASI.. ! YA BASI. Skng itu yg di butuhkan TINDAKAN Aparat penegak HUKUM Menjaga TOLERANSI dan KeBHINEKAAN sesuai UU ID Proses HUKUM.. ! https://t.co/e6nyCPMUQt</p>
<p>Sudah cukup lama kami menunggu tindakan dari para penegak hukum dalam memberantas praktik judi online yang merugikan masyarakat ini.</p> <p>#brantasjudionline</p>

<i>Tokenizing</i>	“apakah”, “ini”, “suatu”, “kebanggaan”, “bagi”, “apparat”, “penegak”, “hukum” “tanya”, “pak”
<i>Filtering</i>	“apakah”, “suatu”, “kebanggaan”, “bagi”, “apparat”, “penegak”, “hukum” “tanya”, “pak”
<i>Stemming</i>	“apa”, “suatu”, “bangga”, “bagi”, “apparat”, “penegak”, “hukum” “tanya”, “pak”

Pada tahap *preprocessing* ini dilakukan beberapa tahapan agar data text yang ada dapat dilakukan untuk proses selanjutnya. Adapun tahapan Proses *Preprocessing* yang dilakukan dapat dilihat dibawah ini :

1. *Cleaning*

Pada tahap ini dilakukan sebuah proses untuk menghilangkan semua karakter dalam sebuah data *crawling* yang bukan merupakan huruf abjad, sehingga dapat mengurangi karakter-karakter yang tidak diinginkan atau tidak bermakna. Karakter-karakter tersebut antara lain seperti angka, tanda pagar (#), tanda @, emoji, atau pun tautan dari situs *web* yang ada di dalam sebuah postingan serta pada proses ini dilakukan penghapusan data duplikat yang kemudian data akan disimpan dalam format csv. Proses *case folding* juga dilakukan pada tahapan ini, pada prosesnya dilakukan pengubahan setiap huruf dalam data text yang ada pada data *crawling* dirubah menjadi huruf kecil (*lowercase*).

Dari hasil data *crawling* yaitu sebanyak 4165 data *tweet* dihasilkan, kemudian dilakukan proses data *cleaning* sehingga data yang di hasilkan sebanyak 3430 data. Dibawah ini pada Gambar 2. merupakan proses data *cleaning* serta pada Gambar 3. merupakan hasil dari proses data *cleaning* yang telah dilakukan.

3.2. Tahap *Preprocessing*

Data yang telah di *crawling* merupakan data mentah yang belum dilakukan proses pembersihan data, sehingga dilakukanlah proses data *Preprocessing* karena data yang diambil merupakan data yang tidak terstruktur. Berikut merupakan contoh tahapan yang dilakukan dalam proses data *preprocessing* dapat dilihat pada Tabel 2.

Tabel 2. Contoh Tahapan Data *Preprocessing*

Tahapan	Hasil
<i>Text</i>	RT @henk_batuwael: @jokerjutex Apakah ini suatu kebanggaan bagi aparat penegak hukum???? Tanya Pak @ListyoSigitP
<i>Cleaning</i>	Apakah ini suatu kebanggaan bagi aparat penegak hukum Tanya Pak
<i>Case Folding</i>	apakah ini suatu kebanggaan bagi aparat penegak hukum tanya pak

```
##Cleaning data
removeURL <- function(x) gsub("http[^\s:]*", "", x)
twitclean <- tm_map(komenc, removeURL)
removeNL <- function(y) gsub("\n", " ", y)
twitclean <- tm_map(twitclean, removeNL)
replacecomma <- function(y) gsub(",", " ", y)
twitclean <- tm_map(twitclean, replacecomma)
removeRT <- function(y) gsub("RT ", "", y)
twitclean <- tm_map(twitclean, removeRT)
removetik2 <- function(y) gsub(":", " ", y)
twitclean <- tm_map(twitclean, removetik2)
removetikikoma <- function(y) gsub(";", " ", y)
twitclean <- tm_map(twitclean, removetikikoma)
removetikik3 <- function(y) gsub("p.", " ", y)
twitclean <- tm_map(twitclean, removetikik3)
removeamp <- function(y) gsub("&", " ", y)
twitclean <- tm_map(twitclean, removeamp)
removeUN <- function(z) gsub("@\w+", "", z)
twitclean <- tm_map(twitclean, removeUN)
remove.all <- function(xy) gsub("[^\s:]*", "", xy)
twitclean <- tm_map(twitclean, remove.all)
twitclean <- tm_map(twitclean, removePunctuation)
twitclean <- tm_map(twitclean, tolower)

dataframe_twitclean<-data.frame(text=unlist(sapply(twitclean, `[]`)),
stringsAsFactors=F)

value_twitclean <-dataframe_twitclean$text
```

Gambar 2. Proses *Cleaning* Data

	Filter
	,"text"
1	1,yg berdosadilarang itu kejahatan perbuatan tdk me...
2	2,apakah ini suatu kebanggaan bagi aparat penegak ...
3	3,ini apaan sih jendral koq ngeri gini itu mereka bisa ...
4	4,yg berdosadilarang itu kejahatan perbuatan tdk me...
5	5,saya tau kasus berbau sara pasti bakalan rumit bua...
6	6,toleransi itu udh diajarkan di keluarga sd smp sma a...
7	7,saya tau kasus berbau sara pasti bakalan rumit bua...
8	8,nanti klo pas diklarifikasi oleh aparat penegak huku...
9	9,perdagangan orang sdh menjadi cerita lama byk ss...
10	10,sudah cukup lama kami menunggu tindakan dari ...
11	11,tanya dong apakah di riau tidak tersedia aparat pe...
12	12,kepada para penegak hukum kami ingin menagih j...
13	13,yg berdosadilarang itu kejahatan perbuatan tdk m...
14	14,apakah ramadhan sama sekali tak mempengaruhi ...
15	15,setuju gw bwang sama lu sama sepak bolah indo a...
16	16,gabisa bro ada aza peradilan cepat sederhana dan...
17	17,bener banget mas brow ngarepin kpk dan aparat p...

Gambar 3. Hasil *Sample* Dari Proses *Cleaning*

2. Tokenizing

Pada tahap *tokenizing* atau tokenisasi maka akan dilakukanlah suatu tindakan memisahkan setiap kata dalam sebuah input string secara individual. Umumnya, setiap kata akan dipisahkan oleh karakter spasi, sehingga pada proses *tokenizing*, pemisahan kata akan mengandalkan karakter spasi yang digunakan. Dari 3430 data text selanjutnya dilakukan proses *tokenizing* sehingga data yang dihasilkan sebanyak 78321 kata. Pada Gambar 4 merupakan proses dari tokenizing

dan pada Gambar 5 merupakan hasil dari proses tokenizing.

```
library(tokenizers)
data_tokenisasi <- tokenize_words(data_stemming_value)
data_tokenisasi <- data.frame(text=unlist(sapply(data_tokenisasi, `[]`)), stringsAsFactors=F)
data_tokenisasi_value <- data_tokenisasi$text
write.csv(data_tokenisasi, file = "4. Data Tokenisasi.csv")
```

Gambar 4. Proses *Tokenizing*

	Filter
	text
106	jahat
107	yg
108	larang
109	ibadah
110	agama
111	layak
112	hukum
113	di
114	dunia
115	akhirat
116	oleh
117	tegak
118	hukum
119	saya
120	tau
121	kasus
122	bau

Gambar 5. *Sample* Hasil Dari Proses *Tokenizing*

3. Filtering

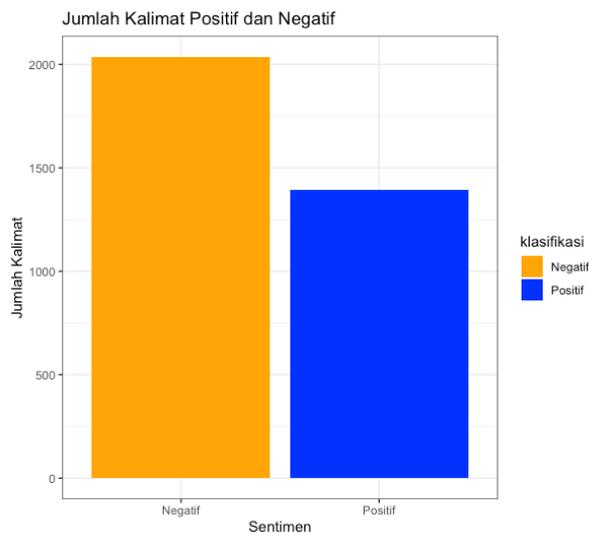
Pada tahap *Filtering* ini dilakukan sebuah proses yang bertujuan untuk menghilangkan kata-kata yang tidak memiliki arti atau biasa disebut sebagai *stopword*. Pada proses *filtering* dihasilkan sebanyak 51872 kata, dibawah ini pada Gambar 6 merupakan proses yang dilakukan dalam tahap *filtering*.

```
library(stopwords)
install.packages("tm")
library(tokenizers)
install.packages("tm")
library(tm)

myStopwords <- readLines("~/Documents/Penegak Hukum Serem 2/stopword_list_id.2.1.txt")
data_stopword <- as.character(data_tokenisasi)
data_stopword <- tokenize_words(data_stopword, stopwords = myStopwords)
data_stopword <- data.frame(text=unlist(sapply(data_stopword, `[]`)), stringsAsFactors=F)
write.csv(data_stopword, "5. Data Stopword.csv")
```

Gambar 6. Proses *Filtering*

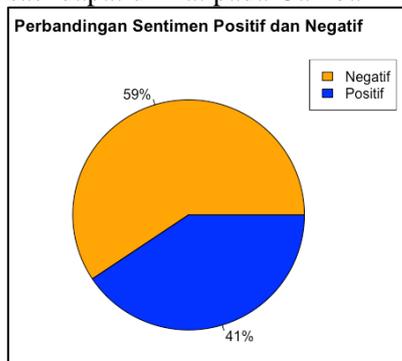
Dari hasil yang proses filtering yang telah dilakukan, maka menghasilkan 10 kata frekuensi tertinggi yang sering muncul, Gambar 7 menunjukkan digram kata frekuensi tertinggi.



Gambar 10. Jumlah Kalimat Positif dan Negatif

3.6. Perbandingan Sentimen

Hasil dari analisis sentimen yang telah dilakukan dengan jumlah data yang dilakukan pengujian adalah sebanyak 4165 data *tweet* yang kemudian di implementasi dengan metode *Naïve Bayes* menghasilkan 59% sentimen negatif dan 41% sentiment positif. Hasil dari visualisasi dapat dilihat pada Gambar 11.



Gambar 11. Perbandingan Sentimen Positif dan Negatif

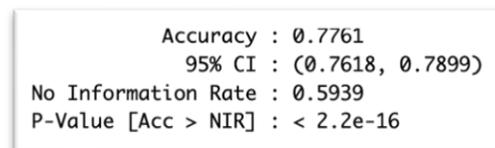
3.7. Penerapan Algoritma Naïve Bayes

Setelah dilakukan proses *crawling* data, *text pre-processing*, serta sentimen analisis, selanjutnya dilakukan tahap terakhir yaitu penerapan algoritma *Naïve Bayes* untuk menentukan nilai akurasi dari sebuah metode yang diterapkan.

Pada tahapan implementasi algoritam *Naïve Bayes* ini, data terlebih dahulu dibagi menjadi data *training* serta data *testing*. Selanjutnya dilakukan prediksi antara model

yang telah dibuat dengan data latih yang kemudian dibandingkan terhadap data uji.

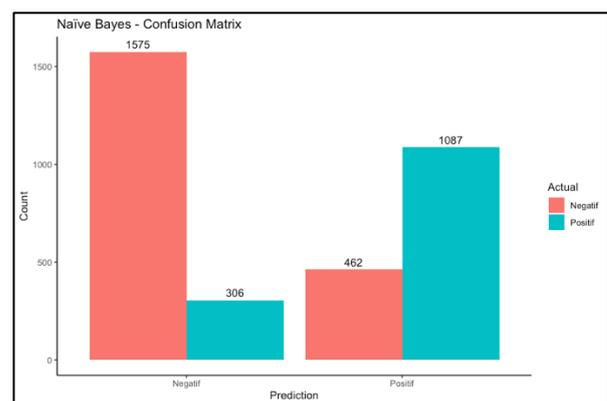
Berdasarkan pengolahan *text* dengan menggunakan algoritma *Naïve Bayes* pada kata penegak hukum, maka Langkah selanjutnya adalah melakukan proses perhitungan untuk melihat nilai akurasi terhadap algoritma yang diterapkan. Dapat lihat pada Gambar 12 merupakan hasil perhitungan dan nilai akurasi sebesar 95%.



Gambar 12. Nilai Akurasi

Pada proses pengujian yang telah dilakukan dengan menggunakan algoritman *Naïve Bayes* maka di peroleh perbandingan terhadap nilai prediksi dan actual terhadap sentimen masyarakat menghasilkan grafik nilai actual lebih tinggi dari pada nilai prediksi, baik itu dengan kategori kalimat positif ataupun dengan kategori kalimat negatif.

Dapat dilihat pada Gambar 13 merupakan grafik perbandingan nilai prediksi dan actual.



Gambar 13. Nilai Prediksi dan Actual

3.8. Analisis Hasil

Berdasarkan hasil yang diperoleh dari situs <https://indikator.co.id> menunjukkan hasil survei pada tanggal 11-17 Agustus 2022 terhadap penegakan hukum di Indonesia terdapat hasil dengan kategori sangat baik sebesar 4,5%, kategori baik sebesar 25%, kategori sedang sebesar 26,5%, kategori sangat buruk sebesar 8,1% dan kategori tidak menjawab sebesar

6,4%. Dengan demikian dapat diartikan menurut data hasil survei dengan rata-rata nilai pada kategori sangat baik dan baik sebesar 29,5% sedangkan rata-rata nilai pada kategori buruk dan sangat buruk adalah sebesar 37,7% dan 31,8 % merupakan termasuk kedalam kategori sedang dan tidak menjawab.

Terkait grafik nilai survei penegakan hukum nasional dapat dilihat pada Gambar 14.



Gambar 14. Survei Penegakan Hukum Nasional

Sumber : [21]

Dari proses analisis sentimen terhadap kepercayaan masyarakat pada proses penegakan hukum di Indonesia, telah dihasilkanlah nilai sebesar 59% dengan kategori sentimen negatif dan 41% dengan kategori sentimen positif. Dari nilai yang diperoleh hal ini menunjukkan bahwa penilaian negatif menduduki nilai tertinggi dengan begitu dapat diartikan bahwa nilai kepercayaan masyarakat berkurang terhadap penegakan hukum di Indonesia. Perbandingan nilai yang muncul antara sentimen positif dan negatif yaitu sebesar 18%, nilai tersebut dapat dikategorikan rendah. Maka, hal ini dapat disebabkan karena beberapa faktor yang terlibat diantaranya, yaitu :

1. Pada survei nasional terbentuk 6 kategori, sedangkan pada analisis sentimen hanya terbentuk 2 kategori.
2. Pada analisis sentimen, kalimat yang seharusnya bernilai positif tetapi dalam proses yang dilakukan oleh mesin berubah menjadi kalimat negatif, begitupun sebaliknya, hal ini tentu saja mempengaruhi dalam pengkategorian

kalimat sentimen sehingga tidak terjadinya akurasi yang tepat dalam pemilihan kalimat sentimen yang bermakna positif dan negatif.

Dari uraian diatas maka perlu dilakukan kembali analisis sentimen dalam tahap kategori yang lebih kompleks sesuai dengan maksud kalimat, agar terbentuk suatu analisis sentimen yang valid dan akurat. Pada proses analisis sentimen menggunakan mesin terkadang kalimat yang semestinya dianggap positif oleh manusia, dapat dianggap negatif oleh mesin dan sebaliknya.

Ketidakkuratan dalam pengkategorian kalimat sentimen ini dapat mempengaruhi hasil analisis sentimen secara keseluruhan. Jika mesin salah mengidentifikasi sentimen dalam sejumlah kalimat, maka keseluruhan hasil analisis sentimen yang dihasilkan juga dapat menjadi tidak akurat. Oleh karena itu, perlu dilakukan perbaikan pada algoritma mesin agar dapat mengidentifikasi sentimen dengan lebih akurat dan dapat memahami konteks dan makna yang lebih dalam dari kalimat.

4. KESIMPULAN

Setelah dilakukannya proses analisis sentimen dengan menggunakan *tools* RStudio terhadap kata penegak hukum pada *Twitter* dengan jumlah data sebanyak 4165 data ulasan, maka diperolehlah sentimen positif sebesar 41% dan sentimen negatif sebesar 59%. Serta dari analisis hasil yang diperoleh terkait perbandingan nilai yang muncul antara sentimen positif dan negatif yaitu sebesar 18%, nilai tersebut bisa dikategorikan rendah karena mempunyai berbagai aspek yang mempengaruhinya yaitu ketidakkuratan sistem dalam pengkategorian kalimat sentimen maka dapat mempengaruhi nilai sentimen yang akan dihasilkan.

Pada nilai akurasi yang diperoleh dalam menerapkan algoritma *Naïve Bayes* menghasilkan nilai sebesar 95%, nilai akurasi yang diperoleh sangat tinggi sehingga tepat untuk diterapkan dalam analisis sentimen pada pada penelitian ini.

DAFTAR PUSTAKA

- [1] H. Moho, "Penegakan Hukum di Indonesia Menurut Aspek Kepastian Hukum, Keadilan dan Kemanfaatan," *Jurnal Warta*, 2019.
- [2] H. Arianto, "Hukum Responsif dan penegakan Hukum di Indonesia," *Lex Jurnalica*, vol. 7, no. 2, 2010.
- [3] M. G. Rahman, S. Tomayahu, F. Syariah, I. Sultan, and A. Gorontalo, "Penegakan Hukum Di Indonesia," 2020. [Online]. Available: <http://journal.iaingorontalo.ac.id/index.php/ah>
- [4] E. Alfian, "Tugas dan Fungsi Kepolisian Untuk Meningkatkan Kepercayaan Publik Terhadap Penegak Hukum," *Legalitas: Jurnal Hukum*, vol. 12, no. 1, p. 27, Jun. 2020, doi: 10.33087/legalitas.v12i1.192.
- [5] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *Jurnal Sains dan Informatika*, vol. 7, no. 2, pp. 155–164, Dec. 2021, doi: 10.34128/jsi.v7i2.331.
- [6] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," 2019.
- [7] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes", [Online]. Available: <http://dev.twitter.com>.
- [8] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and Samudi, "Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine," *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 2, pp. 2502–714, 2020.
- [9] R. Fajar, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 3, no. 1, 2018.
- [10] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2021.
- [11] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," 2022.
- [12] M. Rifki and I. Imelda, "Analisis Sentimen Wacana Kenaikan Harga Tiket Candi Borobudur Menggunakan Multinomial Naïve Bayes," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 2, 2022, doi: 10.33387/jiko.
- [13] S. Lestari *et al.*, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindehan Ibu Kota Negara Indonesia pada Twitter," *JUSIFO (Jurnal Sistem Informasi)*, vol. 8, no. 1, pp. 13–22, 2022.
- [14] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, pp. 157–163, Jan. 2021, doi: 10.30865/mib.v5i1.2604.
- [15] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT*, vol. 7, no. 1, 2021.
- [16] B. Gunawan, H. Sasty Pratiwi, and E. Esyudha Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com
- [17] J. E. Bororing and F. Faeruzah, "Analisis Sentimen Layanan Akademik Menggunakan Metode Naive Bayes Classifier Pada Komentar Mahasiswa,"

- Jurnal Informasi Interaktif*, vol. 5, no. 3, 2020.
- [18] R. Khalida and S. Setiawati, "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *Journal of Information and Information Security (JIFORTY)*, vol. 1, no. 1, 2020, [Online]. Available: <http://ejurnal.ubharajaya.ac.id/index.php/jiforty>
- [19] T. N. Wijaya, R. Indriati, and M. N. Muzaki, "Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3, no. 2, 2021, [Online]. Available: <https://t.co/TIk5mK5bwS>
- [20] D. S. Utami, A. Erfina, and M. A. Id, "Analisis Sentimen Ulasan Terkait UNESCO Global Geopark Di Google Maps dengan Algoritma Naive Bayes," 2022.
- [21] INDIKATOR, "Rilis Hasil Survei Nasional INDIKATOR 'Persepsi Publik terhadap Kasus Sambo: Antara Penegakan Hukum dan Harapan Warga,'" <https://indikator.co.id/rilis-survei-25-agustus-2022/>, JAKARTA PUSAT, Aug. 25, 2022.