

# ANALISIS SENTIMEN MASYARAKAT MENGENAI KASUS KEKUASAAN NARKOBA PADA KOMENTAR YOUTUBE MENGUNAKAN METODE K-NEAREST NEIGHBOURS

Dandy Firmansyah Fitrianto Sutrisno<sup>1</sup>, Sejati Waluyo<sup>2\*</sup>

<sup>1,2\*</sup>Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1</sup>1911500260@student.budiluhur.ac.id, <sup>2\*</sup>sejati.waluyo@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-**Sebagai salah satu media sosial yang populer di masyarakat Indonesia, *platform* Youtube menampilkan berbagai video seperti berita terkini. TvOneNews selalu memberikan dan menayangkan informasi yang sangat *up to date*. Seperti yang diketahui, informasi yang sedang hangat dibahas oleh TvOneNews mengenai maraknya kasus narkoba di Indonesia, maka penulis tertarik untuk melakukan sebuah penelitian analisis sentimen masyarakat pada komentar kanal Youtube TvOneNews dengan video yang berjudul “Polisi Ungkap Jaringan Narkoba Internasional Terbesar Senilai 10 Triliun Rupiah”. Penelitian ini menggunakan metode *K-Nearest Neighbours(KNN)* untuk menentukan sentimen masyarakat *positive* atau *negative*. Sistem yang dibuat menerapkan fitur *crawling* melalui *API* Youtube untuk melakukan pengumpulan *data* komentar, menerapkan *preprocessing*, pembobotan *value* kata dengan *countvectorizer*, evaluasi pengujian akhir dengan metode *confusion matrix*. *Data* yang terkumpul dari proses *crawling* sebanyak 2695 komentar, setelah melalui tahap *preprocessing* menjadi 608 komentar, lalu *data* dibagi menjadi *data train* dan *data test* dengan perbandingan 90:10. Hasil yang didapatkan dalam pengujian akhir menggunakan *confusion matrix*, hasil terbaik dengan nilai  $k=5$  mendapat nilai *accuracy* 85.25%, *precision* 30.0%, *recall* 60.0%, *f1-score* 40.0%. sedangkan nilai pengujian terendah terdapat pada nilai  $k = 11$  yaitu *accuracy* 83.61%, *precision* 10.0%, *recall* 50%, *f1-score* 16.67%. Berdasarkan hasil evaluasi dari sistem yang dibuat, sentimen masyarakat terhadap kasus jaringan narkoba adalah *negative*.

**Kata Kunci:** *Text mining, Youtube API, KNN, Countvectorizer, Confusion matrix*

## ANALYSIS OF PUBLIC SENTIMENT REGARDING THE CASE OF DRUG POWER ON YOUTUBE COMMENTS USING THE K-NEAREST NEIGHBOURS METHOD

**Abstract-** As one of the most popular social media in Indonesian society, the Youtube platform displays various videos such as the latest news. TvOneNews always provides and conveys very up to date information. As is known, information that is being hotly discussed by TvOneNews is regarding the rise of drug cases in Indonesia, so the author is interested in conducting research on public sentiment in the comments on the TvOneNews Youtube channel with a video entitled "Police Reveal the Largest International Drug Network Worth 10 Trillion Rupiah". This research uses the K-Nearest Neighbors (KNN) method to determine positive or negative community sentiment. The system created applies the crawling feature via the Youtube API to collect comment data, apply preprocessing, weight word values with a countvectorizer, evaluate the final test using the confusion matrix method. The data collected from the crawling process was 2695 comments, after going through the preprocessing stage it became 608 comments, then the data was divided into train data and test data with a ratio of 90:10. The results obtained in the final test used a confusion matrix, the best result with a value of  $k=5$  received an accuracy value of 85.25%, precision 30.0%, recall 60.0%, f1-score 40.0%. while the lowest test value is  $k = 11$ , namely accuracy 83.61%, precision 10.0%, recall 50%, f1-score 16.67%. Based on the evaluation results of the system created, public sentiment towards drug network cases is negative.

**Keywords:** *Text mining, Youtube API, KNN, Countvectorizer, Confusion matrix*

## 1. PENDAHULUAN

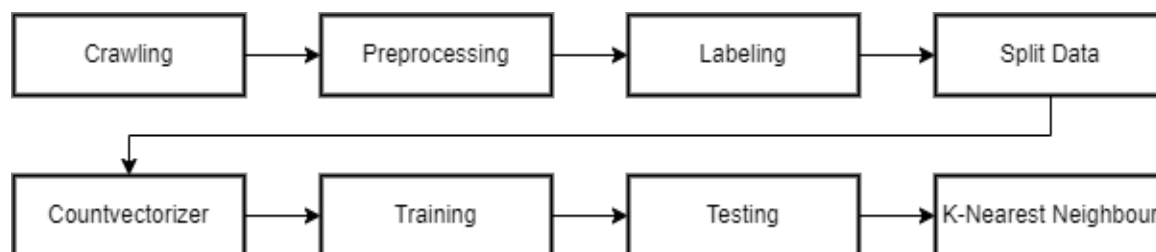
Modernisasi membawa kemajuan teknologi yang pesat. Hanya bermodalkan *smartphone* atau *PC*, membuat masyarakat lebih mudah dalam penggunaannya. Dengan adanya *smartphone* atau *PC* yang terhubung dengan akses internet, masyarakat menjadi lebih aktif berinteraksi di sosial media, salah satunya yaitu Youtube. Salah satu *channel* yang memberikan informasi tentang permasalahan di Indonesia yaitu TvOneNews. TvOneNews membahas kasus jaringan narkoba dengan total komentar 2.695 pada 13 September 2023 pada video yang berjudul “Polisi Ungkap Jaringan Narkoba Internasional Terbesar Senilai 10 Triliun Rupiah”.

Penulis tertarik untuk melakukan sebuah penelitian untuk mendeteksi sentimen masyarakat dikomentar Youtube yang berjudul “Polisi Ungkap Jaringan Narkoba Internasional Terbesar Senilai 10 Triliun Rupiah”. Meskipun penelitian sejenisnya yang dilakukan sebelumnya dengan berbagai macam metode seperti penelitian yang dilakukan oleh Safira dan Hasan pada Januari 2023 yang berjudul “ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER” membahas opini masyarakat terhadap paylater menggunakan metode naive bayes. Penelitian ini menggunakan algoritme *K-Nearest Neighbours (KNN)* untuk menentukan seberapa besar reaksi masyarakat terhadap kasus yang lebih terbaru ini. Kontribusi penelitian ini dilakukan yang bertujuan untuk mengklasifikasikan opini masyarakat pada sosial media Youtube terhadap kekuasaan narkoba tersebut. Metode yang digunakan adalah melakukan analisis sentimen melalui pendekatan *machine learning*, dengan ekstraksi fitur menggunakan *countvectorizer* dengan algoritme *KNN*.

*Dataset* yang digunakan diambil dari komentar video tersebut. Penelitian ini menggunakan algoritme *K-Nearest Neighbours (KNN)* untuk menentukan seberapa besar reaksi masyarakat terhadap kasus jaringan narkoba. Penelitian ini akan menggunakan metode analisis sentimen melalui pendekatan *machine learning*, dengan ekstraksi fitur menggunakan *countvectorizer* dengan algoritme *KNN*.

## 2. METODE PENELITIAN

Dalam merancang aplikasi analisis sentimen, peneliti melakukan beberapa tahapan yang menjadi representasi proses dari tahap awal hingga akhir. Tahapan yang dilakukan dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

Pada tahap pertama dilakukan *crawling* atau mengumpulkan *data* komentar Youtube untuk digunakan sebagai *dataset*, kemudian *data* komentar yang terkumpul disimpan dalam bentuk *file \*.csv*. Tahap *preprocessing* yaitu tahap filtering atau pembersihan *data* komentar dan menghasilkan *data* komentar bersih. Tahap *labeling* memberikan label secara manual dengan dua kategori sentimen yaitu *positive* dan *negative*. Tahap *split data*, *data* akan dibagi menjadi *data train* (90%) dan *data test* (10%). Tahap *Countvectorizer* dilakukan pada *data test* untuk mengetahui frekuensi kata yang muncul berdasarkan *list* kata. *Data train* akan dilatih dan digunakan sebagai model untuk melakukan pengujian terhadap *data test*. *Data test* digunakan untuk menguji akurasi model dengan menerapkan algoritme *KNN* dan *Confusion matrix* untuk mengetahui nilai *accuracy*, *precision*, *recall* dan *f1-score* dan visualisasi.

### 2.1 Text mining

*Text mining* adalah metode untuk memecahkan masalah klasifikasi, *clustering*, pencarian informasi, dan layanan informasi. Perbedaan antara *text mining* dan *data mining* yaitu model yang digunakan, *text mining* menggunakan kumpulan teks yang belum diproses atau terorganisir dengan baik. Model *data mining* diambil dari *database* yang terstruktur[1]. Analisis sentimen merupakan salah satu bidang penelitian *text mining* yang bertujuan untuk mengetahui pendapat atau subjektivitas masyarakat terhadap suatu topik, peristiwa, atau topik pembahasan tertentu. Analisis sentimen adalah proses mengekstraksi, memproses, dan memahami *data* teks tidak terstruktur secara otomatis dan digunakan untuk mengumpulkan opini dan opini tentang frasa opini. *Data* pengujiannya berupa komentar masyarakat yang dilakukan secara daring yang memuat pernyataan emosi berupa senang, sedih, ataupun marah sehingga akan mendapat penilaian *positive* dan *negative*. Youtube membuat para peneliti dapat lebih mendalami dan mempelajari tentang suatu emosi dan opini masyarakat dengan analisis sentimen[2].

### 2.2 Youtube API

Youtube menyediakan layanan yang disebut Youtube *API* untuk pengembang. *API* yang tersedia berupa *API* untuk publik yang berarti pihak Youtube menyediakan akses *data* dan akses layanan kepada publik. Youtube menyediakan Youtube *API* untuk para developer membuat suatu aplikasi yang berinteraksi langsung dengan sumber video yang berada di *platform* Youtube tersebut[3].

### 2.3 Preprocessing

*Preprocessing* merupakan langkah pertama dari proses *text mining*. Rutinitas dan proses untuk mempersiapkan *data* yang akan digunakan dalam fungsi pengambilan *data* sistem *text mining* tercakup dalam langkah ini. Tujuan dari *preprocessing* adalah untuk membuat *data* lebih mudah dan efisien untuk kebutuhan pemrosesan selanjutnya. Selama *preprocessing*, Semua kata dalam dokumen yang dipilih dicocokkan untuk membuat kata unik, dan kata kunci dipilih dari kamus berdasarkan frekuensi kata dalam dokumen tersebut[4]. Berikut adalah penerapan *preprocessing*[5]:

- a. *Cleansing* yaitu proses menghapus simbol atau karakter yang tidak digunakan pada pesan teks.
- b. *Case folding*, yaitu proses merubah huruf kapital menjadi huruf kecil.
- c. *Normalization*, yaitu proses merubah kata singkatan, *slang* dan lainnya menjadi kata yang sesuai dengan KBBI.
- d. *Tokenization*, yaitu proses memisahkan teks per-kata.
- e. *Stopword Removal*, yaitu proses membuang kata yang tidak memiliki arti.
- f. *Stemming*, yaitu proses merubah kata majemuk menjadi kata dasar.

### 2.4 Pelabelan

Pelabelan adalah proses pendefinisian kategori berdasarkan ciri atau ciri yang terdapat pada suatu dokumen atau kalimat. Setiap komentar dalam penelitian ini diberi label positif atau negatif, yang merupakan klasifikasinya. Label positif yang diberikan pada komentar memiliki arti yaitu komentar yang cenderung menerima/mendukung/memuji jaringan narkoba, sedangkan label negatif menunjukkan komentar memiliki arti tidak menerima/mendukung/memuji jaringan narkoba[6].

### 2.5 Pembagian Data

*Data* yang sudah diberi label akan dibagi menjadi dua *data*. Pembagian *data* dilakukan dengan cara membagi *dataset* dengan perbandingan 9:1 yaitu 90% dan 10%. *Data* latihan 90% dan *data* uji 10%[6].

### 2.6 K-Nearest Neighbours

*K-Nearest Neighbor* merupakan sebuah algoritma pembelajaran mesin untuk *classification* objek baru berdasarkan sejumlah k tetangga terdekat. Berdasarkan kategori pada algoritma *K-Nearest Neighbor*, kelas mayoritas akan menjadi kelas baru dari hasil *classification*. Langkah pertama *K-Nearest Neighbor* adalah tentukan nilai k yang akan menjadi jumlah tetangga terdekat. Setelah itu, menghitung jarak objek terhadap *data training* kemudian mengurutkannya secara *ascending* (dari nilai tertinggi ke nilai terendah). tinggi ke rendah (dari nilai yang paling tinggi). Terakhir kumpulkan kelas (klasifikasi tetangga terdekat) dan gunakan kelas tersebut untuk mendapatkan prediksi kelas[7].

### 2.7 Countvectorizer

*Countvectorizer* merupakan metode dasar mengukur kata dengan menghitung berapa kali setiap kata muncul di setiap dokumen, sehingga bisa disebut juga metode penghitungan kasar. Hasil menerapkan metode ini matriks renggang. Meskipun menghitung kemunculan sebuah kata adalah tugas yang tidak rumit, hal ini mungkin tidak signifikan dalam vektor yang diinginkan jika kata tersebut sering muncul di banyak dokumen.[8].

### 2.8 Confusion Matrix

*Confusion matrix* menampilkan pengenalan *dataset* prediksi benar dan salah berdasarkan kemiripannya dengan fakta yang diperoleh. Tabel 1 menyajikan tabel yang memberikan keluaran *confusion matrix*[9].

**Table 1.** Confusion Matrix

Aktual/Prediksi	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	TP	FN
<i>Negative</i>	FP	TN

- a. TP (*True positive*): *Data* opini positif tentang jaringan narkoba yang diprediksi dengan tepat.
- b. TN (*True negative*): *Data* negatif tentang sikap terhadap jaringan narkoba yang diprediksi dengan tepat.
- c. FN (*False negative*): *Data* positif tentang perasaan terhadap jaringan narkoba diharapkan menjadi informasi negatif tentang perasaan terhadap jaringan narkoba.

- d. FP (*False positive*): *Data* negatif dari jaringan narkoba yang diharapkan menjadi *data* positif dari jaringan narkoba.

Performa model dapat dihitung dengan menghitung *accuracy*, *precision*, *recall*, *f1-score*[10]. Berikut adalah rumus perhitungan:

- a. *Accuracy* Persamaan 1 adalah bagian dari prediksi yang benar dan total prediksi yang dibuat oleh pengklasifikasi.

$$Accuracy = TP + TN / TP + TN + FP + FN \quad (1)$$

- b. *Precision* Persamaan 2 memberi tahu kita berapa banyak kasus yang diprediksi dengan benar dan ternyata *positive*. Metrik ini menentukan apakah model tersebut dapat diandalkan atau tidak.

$$Precision = TP / TP + FP \quad (2)$$

- c. *Recall* Persamaan 3 menunjukkan jumlah kasus yang benar-benar *positive* yang dapat diprediksi dengan benar menggunakan model tersebut.

$$Recall = TP / TP + FN \quad (3)$$

- d. *F1-score* Persamaan 4 memberi kita gagasan gabungan tentang metrik presisi dan perolehan. Artinya ketika kita mencoba mengupgrade nilai presisi *Recall* turun dan sebaliknya.

$$F1-score = 2 * Precision * Recall / Precision + Recall \quad (4)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 *Crawling*

Tahap *crawling* mengumpulkan *data* Youtube dan berfokus pada komentar terkait jaringan narkoba internasional yang sedang dibahas oleh *channel* TvOneNews. Proses pengambilan *data* melalui akses Youtube *API* dari *channel* TvOneNews dengan cara input *url* video dengan judul “Polisi Ungkap Jaringan Narkoba Internasional Terbesar Senilai 10 Triliun Rupiah | tvOne” terkumpul *data* komentar sebanyak 2695 *data* lalu *data* diunduh untuk dilakukan tahap selanjutnya yaitu *preprocessing*.

#### 3.2 *Preprocessing*

Setelah *data crawling* sudah diunduh, kemudian *data* diinput lalu akan dilakukan tahap *preprocessing*. Tahap *preprocessing* akan menyiapkan *data* dengan cara *case folding*, *cleansing*, *tokenizing*, *stop words*, *normalization*, *stemming*, *token filtering*, *words count* seperti yang sudah dijelaskan pada tahap *preprocessing*. Kemudian hasil *preprocessing* dapat diunduh untuk dapat diberikan label secara manual pada *microsoft excel* atau dapat klik *button* label *comments* untuk memberikan label secara manual pada aplikasi yang dibuat. Setelah tahap *preprocessing*, *data* hasil *preprocessing* yang diunduh berjumlah 608 dari sebelumnya berjumlah 2695.

#### 3.3 *Split*

Setelah tahap *preprocessing* selesai dan *data* diberi label, *data* dapat dimuat dan pembagian *data* dapat dilakukan. Pada langkah *split data* dibagi dua yaitu *data train* dan *data test* dengan cara memasukkan *data* berlabel secara manual, kemudian *data* tersebut dibagi menjadi dua persentase yaitu 90% *data train* dan 10% *data test*. Setelah pemisahan selesai, kedua *data* dapat diunduh untuk diproses lebih lanjut. Dari total 608 *data*, terbagi menjadi total 547 *data train* dan total 61 *data test*.

#### 3.4 *Countvectorizer*

Pertama akan dilakukan pembuatan *list* kata untuk menghitung jumlah vektor Tabel 4 pada setiap *data* Tabel 2. *Data* Tabel 2 akan dipisahkan kata per-kata dan kata yang memiliki duplikasi tidak digunakan lalu ditampung oleh *table list* kata Tabel 3.

**Table 2 Data Train**

No	Nama Akun	Komentar	Label
1	@MangaraSihaloho-ug5zv	kira uang mana uang kontong pribadi kas negara narkoba salah bisnis oknum	<i>negative</i>
2	@riyosori6931	bnn tingkat alat alat canggih cepat narkoba mana jalan cepat tangkap	<i>negative</i>
3	@user-fw1pq7hi7o	apresiasi jajar polisi mhn awas hasil barang sita biar tidak jual oknum tidak tanggung	<i>positive</i>

**Table 3 List Kata**

<i>List Kata</i>
['alat', 'apresiasi', 'awas', 'barang', 'biar', 'bisnis', 'bnn', 'canggih', 'cepat', 'hasil', 'jajar', 'jalan', 'jual', 'kas', 'kira', 'kontong', 'mana', 'mhn', 'narkoba', 'negara', 'oknum', 'polisi', 'pribadi', 'salah', 'sita', 'tanggung', 'tangkap', 'tidak', 'tingkat', 'uang']

**Table 4 Vektor Data Train**

<i>List Kata</i>	<i>Train 1</i>	<i>Train 2</i>	<i>Train 3</i>
alat	0	2	0
apresiasi	0	0	1
awas	0	0	1
barang	0	0	1
biar	0	0	1
bisnis	1	0	0
bnn	0	1	0
canggih	0	1	0
cepat	0	2	0
hasil	0	0	1
jajar	0	0	1
jalan	0	1	0
jual	0	0	1
kas	1	0	0
kira	1	0	0
kontong	1	0	0
mana	1	1	0
mhn	0	0	1
narkoba	1	1	0
negara	1	0	0
oknum	1	0	1
polisi	0	0	1
pribadi	1	0	0
salah	1	0	0
sita	0	0	1
tanggung	0	0	1
tangkap	0	1	0
tidak	0	0	2
tingkat	0	1	0
uang	2	0	0

Selanjutnya akan dilakukan perhitungan frekuensi kemunculan kata muncul pada *data test* Tabel 5 berdasarkan *list* kata yang sebelumnya sudah dibuat. Sebelum proses perhitungan vektor Tabel 6, *data test* akan diubah kata per-kata.

**Table 5 Data Test**

<i>Data</i>	<i>Komentar</i>	<i>List</i>
<i>Data Test</i>	uang sita guna beli alat tempur alustista moderen biar nkri tidak remeh negara uang sitaan korupsi narkoba oknum pengkhianat nkri beli alat perang alustista canggih bagi bikin bonus jabat tidak barokah bro	['alat', 'aset', 'barokah', 'beli', 'bonus', 'bro', 'canggih', 'guna', 'hilangkan', 'jabat', 'kata', 'korusi', 'modern', 'narkoba', 'negara', 'nkri', 'oknum', 'pengkhianat', 'perang', 'remeh', 'sebagai', 'sekutu', 'sita', 'uang']

**Table 6 Vektor Data Test Dengan List Kata Data Train**

<i>List Kata</i>	<i>Data Test</i>
alat	3
apresiasi	0
awas	0
barang	0
biar	1
bisnis	0
bnn	0
canggih	1
cepat	0

List Kata	Data Test
hasil	0
jajar	0
jalan	0
jual	0
kas	0
kira	0
kontong	0
mana	0
mhn	0
narkoba	1
negara	1
oknum	1
polisi	0
pribadi	0
salah	0
sita	1
tanggung	0
tangkap	0
tidak	2
tingkat	0
uang	2

### 3.5 K-Nearest Neighbours

Setelah tahap *countvectorizer* dilakukan, akan dilakukan perhitungan jarak antar *data* dengan perhitungan *euclidean distance* untuk mencari tetangga terdekat. Setelah tetangga terdekat ditentukan akan dihitung jumlah label yang sering muncul untuk dijadikan hasil prediksi sentimen pada *data test*. Berikut Tabel 7 adalah contoh perhitungan jarak dengan *euclidean distance*.

**Table 7** Representasi Vektor *Data Train* dan *Data Test*

Data	Vektor
<i>Train 1</i>	[0,0,0,0,0,1,0,0,0,0,0,0,0,1,1,1,1,0,1,1,1,0,0,1,1,0,0,0,0,2]
<i>Train 2</i>	[2,0,0,0,0,0,1,1,2,0,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,1,0,1,0]
<i>Train 3</i>	[0,1,1,1,1,0,0,0,0,1,1,0,1,0,0,0,0,1,0,0,1,1,0,0,1,1,0,2,0,0]
<i>Data Test</i>	[3,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,0,2,0,2]

Jarak antara *Data Test* dengan *Train 1*:

$$\begin{aligned}
 & \sqrt{(3-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (1-0)^2 + (0-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-0)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2 + (2-0)^2 + (0-0)^2 + (2-2)^2} \\
 & = 9+0+0+0+1+1+0+1+0+0+0+0+0+1+1+1+1+0+0+0+0+0+1+1+1+0+0+0+0+1+1+1+0+0+4+0+0 \\
 & = \sqrt{23} \\
 & = 4.80
 \end{aligned}$$

**Table 8** Hasil Euclidean Distance

Data Train	Jarak Data Test	Label
<i>Train 1</i>	4.80	Negative
<i>Train 2</i>	4.69	Negative
<i>Train 3</i>	5.00	Positive

Berdasarkan hasil perhitungan *euclidean distance* Tabel 8, label yang paling sering muncul pada tetangga terdekat akan menjadi *predicted* label. Pada perhitungan ini tetangga terdekat memiliki dua label *negative* dan satu label *positive*, maka *data test* akan mendapat *predicted* label *negative*.

### 3.6 Pengujian

Dalam pengujian ini *data* yang digunakan adalah 547 *data train* dan 61 *data test*. *Data train* memiliki jumlah sentimen *negative* 455 dan sentimen *positive* 92. *Data test* memiliki jumlah sentimen *negative* 51 dan sentimen *positive* 10. Berikut adalah beberapa *data* hasil hasil *classification K-Nearest Neighbours* pada tabel 9.

**Table 9** Hasil *Classification* K-Nearest Neighbours

Nomor	Actual Label	Predicted Label	Komentar Testing
1	negative	negative	besar loloss operasi senilai puluhan lagi ratus jatim jateng jabar loss narkobaa
2	negative	negative	narkoba hukum mati alasa rusak generasi muda koruptor nga hukum mati alasa nga tau koruptor makan duit rakyat rakyat lena butuh hidup anak
3	positive	positive	polri semangat apresiasi jangan blm apa dibully dsb ntar loyo berantas narkoba
....	....	....	....
61	negative	negative	sedih negri bikin hukum bergihak bandsr narkoba tidak wujud hukum mati peedar bandar salah kelemahan negara tidak tereujud aman generasi muda

Kolom *Actual Label* adalah label yang diperoleh dari proses pelabelan manual oleh pakar, sementara kolom *Predicted Label* adalah label yang diperoleh dari *classification* menggunakan *K-Nearest Neighbours*. Kemudian *confusion matrix* menyatakan semua *data test*. Berikut hasil dari *confusion matrix* tampil pada tabel 10.

**Table 10** Hasil Confusion Matrix

Actual / Predicted	Positive	Negative
Positive	3	7
Negative	2	49

Berdasarkan tabel 10 mendapatkan jumlah sentimen *actual* dan *predicted*, nilai TP adalah 3, FP adalah 7, TN adalah 49, FN adalah 2. Maka perolehan dari tabel 10 akan dihitung *accuracy*, *precision*, *recall*, *f1-score*. Berikut adalah perhitungannya dapat dilihat pada tabel 11 perhitungan dengan nilai k adalah 5.

**Table 11** Hasil Accuracy, Precision, Recall, F1-Score

<b>Accuracy</b>	85.25%
<b>Precision</b>	30.0%
<b>Recall</b>	60.0%
<b>F1-score</b>	40.0%

Pengujian dilakukan dengan variasi nilai k agar terlihat perbedaan antara setiap nilai k yang diuji dan terlihat nilai k yang menghasilkan nilai lebih baik, berikut hasil pengujian pada tabel 12.

**Table 12** Variasi Pengujian

Metric	K = 3	K = 5	K = 7	K = 9	K = 11
<b>Accuracy</b>	83.61%	85.25%	83.61%	85.25%	83.61%
<b>Precision</b>	30.0%	30.0%	20.0%	20.0%	10.0%
<b>Recall</b>	50.0%	60.0%	50.0%	66.67%	50.0%
<b>F1-score</b>	37.5%	40.0%	28.57%	30.77%	16.67%

Berdasarkan tabel 12, diketahui nilai pengujian tertinggi terdapat pada nilai k = 5 yaitu *accuracy* 85.25%, *precision* 30.0%, *recall* 60.0%, *f1-score* 40.0%. sedangkan nilai pengujian terendah terdapat pada nilai k = 11 yaitu *accuracy* 83.61%, *precision* 10.0%, *recall* 50%, *f1-score* 16.67%. Dari hasil pengujian analisis sentimen dengan nilai k = 5 terhadap 61 *data test* sentimen *negative* sebesar 91.8% dan sentimen *postive* sebesar 8.2% membuktikan sentimen masyarakat terhadap kasus jaringan narkoba adalah negatif.

#### 4. KESIMPULAN

Sebagaimana hasil evaluasi dari sistem untuk analisis sentimen masyarakat terhadap kasus jaringan narkoba yang dibahas pada kanal Youtube TvOneNews dapat disimpulkan bahwa sentimen masyarakat mayoritas negatif dalam menanggapi kasus jaringan narkoba. Berdasarkan sistem aplikasi analisis sentimen terhadap kasus jaringan narkoba, dengan *data train* berjumlah 584 dan *data test* 62 menghasilkan sentimen *negative* 91.8%. Penggunaan *countvectorizer* dan metode *K-Nearest Neighbours* untuk melakukan analisis sentimen dapat bekerja dengan baik. Pengujian dengan nilai terbaik didapatkan dengan nilai k=5 yaitu *accuracy* 85.25%, *precision* 30.0%, *recall* 60.0%, *f1-score* 40.0%.

## DAFTAR PUSTAKA

- [1] M. Afdal and L. R. Elita, “Penerapan *Text Mining* Pada Aplikasi Tokopedia Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 8, no. 1, pp. 78–87, 2022.
- [2] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [3] E. Y. S. Sihombing, Tibyani, and B. T. Hanggara, “Pemanfaatan API Youtube dalam Pengembangan Aplikasi Portal Video Penangkaran Kenari untuk Peternak Kenari Berbasis Android,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 7, pp. 2067–2074, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] R. A. Saputra, “Analisis Sentimen Pada Media Sosial Menggunakan Metode K-Nearest Neighbor (K-NN),” *Teknologipintar.org*, vol. 2, no. 8, pp. 1–8, 2022.
- [5] H. A. N. Syifa, A. Nugroho, and R. Firliana, “Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter,” *Jurnal Ilmiah Informatika (JIF)*, vol. 11, no. 1, pp. 54–62, 2023.
- [6] P. Mus and Painem, “Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,” *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, vol. 2, no. 2, pp. 311–319, 2021.
- [7] I. Irawaty, R. Andreswari, and D. Pramesti, “Vectorizer Comparison for Sentiment Analysis on Social Media Youtube: A Case Study,” *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, pp. 69–74, Jan. 2020, doi: 10.1109/IC2IE50715.2020.9274650.
- [8] A. Averina, H. Hadi, and J. Siswanto, “Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik,” *Teknika*, vol. 11, no. 2, pp. 123–128, Jan. 2022, doi: 10.34148/teknika.v11i2.461.
- [9] M. Lestandy, A. Abdurrahim, and L. Syafa’ah, “Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, Jan. 2021, doi: 10.29207/resti.v5i4.3308.
- [10] H. F. Fadli and A. F. Hidayatullah, “Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM,” *Automata*, vol. 2, no. 1, 2021.