

## Implementasi Algoritma Klasifikasi Terhadap Tweet Pornografi Kaum Homoseksual Pada Twitter

Taopik Hidayat<sup>1</sup>; Rangga Pebrianto<sup>2</sup>; Risca Lusiana Pratiwi<sup>3</sup>; Windu Gata<sup>4</sup>; Daniati Uki Eka Saputri<sup>5</sup>

<sup>1,2,4,5</sup>Magister Ilmu Komputer, STMIK Nusa Mandiri

<sup>3</sup>Sistem Informasi, STMIK Nusa Mandiri

<sup>1</sup>14002393@nusamandiri.ac.id, <sup>2</sup>14002396@nusamandiri.ac.id, <sup>3</sup>risca.ral@nusamandiri.ac.id,

<sup>4</sup>windu@nusamandiri.ac.id, <sup>5</sup>daniati.due@nusamandiri.ac.id

**Abstract:** Twitter is one of the social media with the number of users who reach millions of users. The number of Twitter users in 2019 increased by 17 percent in 2018 to 145 million users with a variety of good both positive and bad. The negative impacts that occur such as the spread of status, images, and videos that affect pornography especially among freedom groups. Homosexuals are sexually oriented people who like the same sex that occurs in men, the rejection often experienced by men makes one of the reasons intellectuals use Twitter social media to show their personal relationships, open to each other, socializing with same sex, looking for conversation, to become a place to find a partner. The purpose of this study is to determine the positive and negative sentiments to determine the level of accuracy of intellectual pornography tweets in Indonesia from data taken from Twitter tweets by using the TF-IDF and k-NN methods. The results of this study get an accuracy value of 88.25% containing pornography and the remaining 11.75% not containing pornography will contain news, news, and other information.

**Keywords:** homosexual, sentiment analysis, twitter

**Abstrak:** Twitter merupakan salah satu media sosial dengan jumlah pengguna mencapai jutaan pengguna. Jumlah pengguna Twitter pada tahun 2019 dicatat meningkat 17 persendari tahun 2018 menjadi 145 juta pengguna dengan berbagai dampak baik dampak positif maupun dampak negatif. Dampak negatif yang ditimbulkannya seperti penyebaran status, gambar, dan video yang bersifat pornografi khususnya di kalangan kaum homoseksual. Homoseksual merupakan orang yang berorientasi seksual sebagai penyuka sesama jenis yang terjadi pada kaum pria, Penolakan yang sering dialami kaum homoseksual men-jadikan salah satu alasan kaum homoseksual menggunakan media sosial Twitter untuk menunjukkan identitas diri mereka, saling terbuka, bersosialisasi dengan sesama jenis, mencari penghasilan, hingga menjadi ajang pencarian pasangan. Tujuan dari penelitian ini adalah untuk mengetahui sentimen positif dan negatif untuk mengetahui tingkat akurasi terhadap tweet pornografi kaum homoseksual di Indonesia dari data yang diambil dari tweet Twitter dengan menggunakan metode TF-IDF dan k-NN. Hasil penelitian ini mendapatkan nilai accuracy sebesar 88,25% mengandung unsur pornografi dan sisanya sebesar 11,75 tidak mengandung unsur pornografi akan tetapi berisi iklan, berita, dan informasi lainnya.

**Kata kunci:** homoseksual, sentimen analisis, twitter



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2019 by author and IJSE-Indonesian Journal on Software Engineering.

### A. PENDAHULUAN

Kemajuan era globalisasi saat ini memberi dampak positif dan negatif diberbagai bidang. Namun kehadiran teknologi juga menjadi salah satu sarana untuk terjadinya hal-hal negatif yang berdampak besar seperti penyebaran status, gambar, dan video yang bersifat pornografi di media social (Tekno.kompas.com, 2019). Berita dalam negeri tentang pornografi yang menarik perhatian banyak pihak di tahun 2020 yaitu pornografi pada kaum homoseksual. Beberapa berita tersebut diantaranya, Seorang pria yang ditangkap karena mengunggah konten pornografi untuk menarik pelanggan yang mau dipijat dan tidur dengannya di Banyuwangi (News.detik.com, 2019).

Seorang pria dibunuh karena menolak berhubungan intim oleh pria homoseksual (News.detik.com, 2020).

Kaum homoseksual merupakan kelompok orang yang berorientasi seksual sebagai penyuka sesama jenis yang terjadi pada kaum pria yang dapat disebabkan oleh beberapa faktor seperti penyimpangan genetika, lingkungan atau karena adanya trauma dalam hubungan seksualitas sebelumnya (Pudjajana & Manongga, 2018). Penolakan yang sering dialami kaum homoseksual yang menjadikan alasan kaum tersebut menggunakan media sosial guna mengekspos identitas diri mereka..

Beberapa media sosial yang dikhususkan untuk kaum homoseksual yaitu *Grindr*, *Blued*, dan lain-lain. Namun dalam penelitian ini, twitter digunakan untuk mendapatkan data dari kaum homoseksual dengan menggunakan kata khusus. Data dari *tweet Twitter* yang merupakan *text mining* akan dilakukan sentimen analisis dan pengklasifikasian menggunakan metode TF-IDF dan k-NN (Septian et al., 2019).

Twitter merupakan salah satu media sosial yang sangat banyak penggunaannya dengan jumlah mencapai jutaan pengguna. Jumlah pengguna Twitter pada tahun 2019 dicatat meningkat 17 persen dari tahun 2018 menjadi 145 juta pengguna (Tekno.kompas.com, 2019). Penggunaan media social twitter dalam penelitian dikarenakan proteksi terhadap tweet, video, dan gambar yang mengandung unsur pornografi masih dapat diakses dengan sangat mudah apabila ditelusuri secara khusus.

Beberapa penelitian terkait yang membahas sentimen analisis adalah penelitian yang dilakukan oleh Andre Maureen Pudjajana pada tahun 2018 melakukan penelitian sentiment analis terhadap tweet pornografi kaum homoeksual dengan menggunakan lima kata khusus yang menunjukkan bahwa nilai sentimen negatif sangat tinggi dengan total rata-rata adalah 68.4% dengan total rata-rata pengujian Naïve Bayes sebesar 87.48%, dan k-NN sebesar 85.40% (Pudjajana & Manongga, 2018). Suwanda Aditya Saputra dan kawan-kawan yang melakukan sentiment analis terhadap E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes dengan menggunakan PSO di tahun 2017 dengan hasil *cross validation* NB tanpa penambahan PSO adalah nilai *accuracy* 82.30% dan nilai AUC 0.780. Sedangkan dengan penambahan PSO adalah nilai *accuracy* 83.60% dan nilai AUC 0.801% (Aputra et al., 2019). Andi Nurul Hidayat melakukan sentiment analis terhadap wacana politik pada media masa online menggunakan algoritma support vector machine dan naive bayes pada tahun 2015 dengan menghasilkan rata-rata akurasisesbesar 59,98 % dannilaitertinggiakurasisesbesar SVM 90,50% (Hidayat, 2015). Sigit Kurniawan dan kawan-kawan yang melakukan Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online pada tahun 2017 dengan hasil akurasi 78.40% dan AUC 0.850. Hasil dari penelitian ini adalah terdapatnya keefektifan algoritma dalam mengklasifikasikan komentar pada media berita online terhadap tokoh-tokoh politik(Kurniawan et al., 2019). Penelitian yang dilakukan oleh Anita Novantirani yang meneliti tentang sentiment analis terhadap pengguna jasa transportasi umum darat dalam kota dengan menggunakan metode SVM pada twitter tahun 2015 dengan mendapatkan hasil nilai akurasi sebesar 78,12%. Jumlah dan perbandingan data uji dan data latih adalah variabel yang paling berpengaruh dalam sentimen analisis ini (Novantirani et al., 2015). Didik Garbian dan kawan-kawan yang melakukan sentimen analisis terhadap jasa ojek online menggunakan metode naïve bayes pada tahun 2015 dengan hasil yang didapatkan dari akurasi naïve bayes memperoleh ketepatan 80% (Nugroho et al., 2015). Penelitian yang dilakukan oleh Eko Yulian mengenai Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung pada tahun 2018 dengan hasil bahwa terdapat banyak tweet pengguna Twitter kota bandung yang terindikasi mengandung unsur LGBT secara umum masih berkaitan dengan unsur religi yang ditandai dengan kemunculan kata agama yang sangat sering (Yulian, 2018). Nurirwan dan kawan-kawan yang melakukan sentiment analis terhadap Presiden Jokowi dengan preprocessing normalisasi dan stemming menggunakan metode naive bayes dan svm pada tahun 2015 dengan menghasilkan akurasi normalisasi dan stemming pada data sebesar 89,2655% dan yang dinormalisasi saja sebesar 88,7006% menggunakan metode SVM (Saputra et al., 2015). Jeremy Andre, Tresna Maulana dan Aryo yang melakukan penelitian tentang Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan TF-IDF dan k-NN di tahun 2019 pada nilai k=23 dengan hasil akurasi yang optimal sebesar 79.99% (Nugroho et al., 2015). Dan penelitian yang dilakukan oleh Raflialdy dan kawan-kawan tentang Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes

pada tahun 2016 dengan hasil akurasi naïve bayes pada data uji negative memperoleh ketepatan 76% sedangkan untuk data uji positif memperoleh 83% (Raksanagara et al., 2016)

Berbeda dengan penelitian yang telah dilakukan sebelumnya, tujuan dari penelitian ini adalah untuk mengetahui sentimen positif dan negatif tujuannya untuk mengetahui tingkat akurasi terhadap tweet pornografi kaum homoseksual dengan menggunakan metode TF-IDF dan metode k-NN.

## B. TINJAUAN PUSTAKA

### 1. Sentimen Analisis

Sentimen analisis adalah proses analisa dari berbagai kalimat berupa pandangan atau opini sehingga dihasilkan kesimpulan dari berbagai opini yang ada dalam suatu kalimat. Sentimen analisis biasanya digunakan untuk mengolah saran dari pengguna dengan memberikan label pada suatu kalimat menjadi positif, negatif atau netral (Setiowati & Helen, 2018).

### 2. Metode TF-IDF

Metode TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode untuk mencari kata yang hampir sama dengan kategori yang tersedia berdasarkan kepada bobot suatu kata yang ditentukan sebelumnya (Saputro et al., 2017).

Metode ini akan menghitung bobot setiap token  $t$  di dokumen  $d$  dengan rumus:

$$W_{dt} = tf_{dt} * IDF_t$$

Dimana:

$d$  : dokumen ke- $d$

$t$  : kata ke- $t$  dari kata kunci

$W$  : bobot dokumen ke- $d$  terhadap kata ke- $t$

$tf$  : banyaknya kata yang dicari pada sebuah dokumen

IDF : Inversed Document Frequency

Nilai IDF didapatkan dari  $IDF = \log_2 (D/df)$

Dimana :

$D$  : total dokumen

$df$  : banyak dokumen yang mengandung kata yang dicari

Setelah bobot ( $W$ ) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai  $W$ , semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

### 3. Metode k-NN

K-NN (*K-Nearest Neighbors*) adalah metode pengklasifikasi objek berdasarkan data dengan berdasarkan kepada jarak terdekat atau kemiripan pada objek tertentu (Pudjajana & Manongga, 2018). Metode ini tidak terlalu membutuhkan sumber daya komputer yang besar dan mudah diimplementasikan (Saputro et al., 2017).

Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak euclidean. Berikut beberapa formula yang digunakan dalam algoritma KNN:

Jarak Euclidean adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

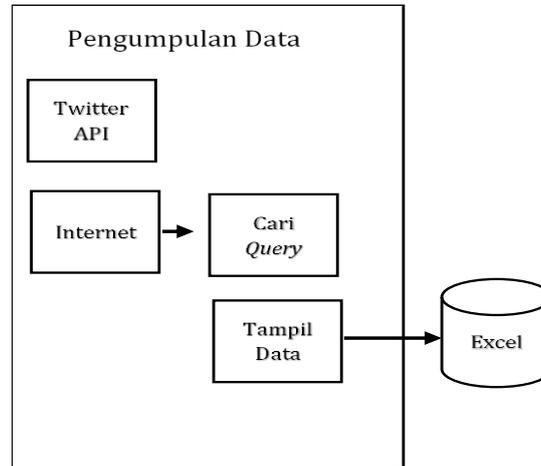
## C. METODE PENELITIAN

### 1. Penentuan Query

Penelitian ini dimulai dengan melakukan analisa dan menentukan *query* yang berkaitan dengan homoseksual yang paling sering muncul pada pesan dalam *Twitter* dikenal dengan sebutan *tweet* (Nugroho et al., 2015). Dengan ditentukannya *Query* pada *Twitter* akan membantu dalam pengumpulann data *tweet* sehingga akan lebih terarah (Aaputra et al., 2019). *Query* yang dipilih dalam penelitian ini yaitu *#gayindonesia*.

## 2. Pengumpulan Data

Pengumpulan data dilakukan berdasarkan *query Twitter* yang telah ditentukan. Data didapatkan dengan menggunakan *Twitter API* dan menggunakan Rapidminer Studio versi 9.6. Data-data tersebut disimpan ke Microsoft Excel 2016 dimana *limit data* terdiri dari 1000 data. Gambar 1 menunjukkan proses pengumpulan data menggunakan Rapidminer Studio.



Gambar 1. Pengumpulan data dengan Rapid Miner Studio

(Sumber: Pudjajana, 2018)

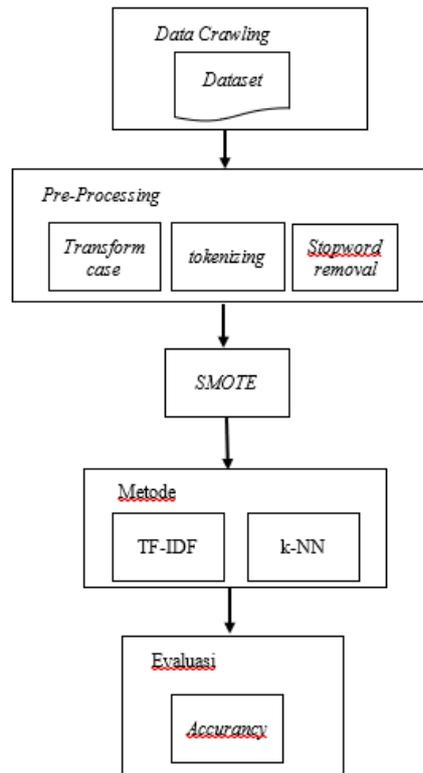
Pada Gambar 1 menjelaskan proses pengumpulan data yang dimulai dengan mengatur *Twitter API* pada Rapid Miner Studio dengan bantuan koneksi internet. Kemudian *Twitter API* harus dikoneksikan dengan *twitter* dengan cara masuk menggunakan akun *Twitter*. Langkah berikutnya adalah mengidentifikasi data dengan cara memasukkan *query* dan memilih bahasa Indonesia lalu dijalankan dan data akan tampil sesuai dengan *query* yang telah ditentukan (Pudjajana & Manongga, 2018). Data pada Rapid miner Studio disimpan ke Microsoft Excel untuk dilakukan proses selanjutnya.

## 3. Pre-processing Data

Pada proses pre-processing data, tweet mentah akan diproses dengan beberapa tahap, yaitu *transform case*, *tokenizing*, dan *stopword removal* (Pudjajana & Manongga, 2018; Anggraini & Suroyo, 2019; Aputra et al., 2019; Saputra et al., 2015). Berikut adalah penjabaran dari proses *pre-processing* data:

- a. *Transform case* adalah mengubah karakter menjadi ukuran yang sama (C & Lukito, 2017). Misalnya 'Mungkin' menjadi 'mungkin'. Hal ini dilakukan agar jika dilakukan proses ke dalam model klasifikasi terdapat kesamaan jenis huruf dan tidak terjadi kesalahan dalam proses *tokenizing*.
- b. *Tokenizing* yaitu semua kata yang ada di dalam tiap data dikumpulkan dan dihilangkan karakter yang tidak diperlukan. Karakter yang dihilangkan adalah karakter HTML, retweet (RT), *emoticon*, *hashtag* (#), *username* (@username), url (<http://website.com>).
- c. *Stopword removal* adalah proses menghilangkan kata-kata yang tidak penting dalam kalimat namun sering muncul dan tidak memiliki pengaruh apapun. Misalnya kata yang penunjuk waktu, kata penghubung dan kata tanya (Hidayat, 2015).

Berdasarkan penjelasan diatas, maka proses pre-processing data dapat digambarkan seperti pada Gambar 2 dibawah ini.



Gambar 2. Model Proses *Pre-Processing* Data Termodifikasi  
(Sumber: Taopik, 2020)

#### D. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengetahui sentimen analisis dari tweet homoseksual di Twitter dan membandingkan hasil akurasi antara metode TF-IDF dan k-NN sesuai dengan Gambar 1. Selain itu, hasil penelitian ini juga mengingatkan masyarakat Indonesia untuk menggunakan media sosial khususnya Twitter secara tepat, bijaksana, dan bernilai positif.

##### 1. Proses Pengambilan Data

Proses crawling data dilakukan dengan menggunakan operator Search Twitter dengan cara membuat koneksi antara RapidMiner dengan Twitter API. Apabila sudah terkoneksi, kemudian masukkan query #GayIndonesia dengan result type recent or popuar yang artinya data yang akan diambil adalah data yang terbaru dan poluper dengan jumlah limit data yang diambil sebanyak 1000 data tweet. Dari 1000 data tweet tersebut dilakukan data cleansing untuk menghilangkan duplikasi data dan menghasilkan 615 data tweet yang telah diberikan label.

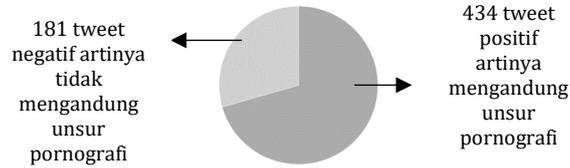
Penelitian ini melakukan sentimen analisis dengan menguji data dengan jumlah 615 tweet yang telah diberi label positif apabila tweet tersebut mengandung unsur pornografi dan label negatif apabila tweet tersebut tidak mengandung unsur pornografi. Data diharuskan mempunyai label dikarenakan dalam penelitian analisis sentimen dengan metode k-NN ini merupakan supervised learning.

Berikut adalah proses pengambilan data (*crawling data*) yang diambil dari Twitter menggunakan Tools RapidMiner Studio versi 9.6.



Gambar 3. Proses Pengambilan Data

Diketahui Gambar 3 adalah gambar proses pengambilan data (*crawling data*) dari Twitter dengan menggunakan fitur *Search Twitter* dengan pengambilan data maksimal 1000 data *twitter*.

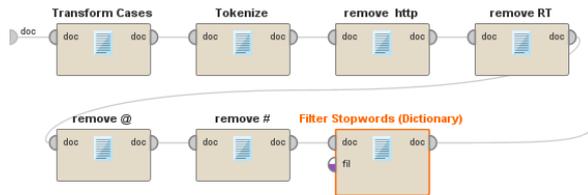


Gambar 4. Diagram *Pie Diagram* Sentimen Positif dan Negatif Dari Dataset

Berdasarkan Gambar 4, Dari 615 tweet terdapat 434 tweet positif mengandung unsur pornografi dan 181 tweet tidak mengandung unsur pornografi, sehingga dapat disimpulkan bahwa dari data yang diambil dari #GayIndonesia ini cenderung mengandung unsur pornografi.

**2. Proses *Pre-Processing* Data**

Tahap selanjutnya adalah melakukan *pre-processing* data yang telah didapat agar dapat diolah dengan tepat pada saat melakukan permodelan. Tahap *pre-processing* meliputi kegiatan membangun data dan juga membersihkan data agar dapat diolah pada tahap selanjutnya. Berikut tahap *pre-processing* data:



Gambar 5. Proses *Pre-Processing* Data

Berdasarkan Gambar 5, dapat diketahui bahwa proses *pre-processing* terdapat operator *Transform case*, *Tokenize* (*remove HTTP*, *remove RT*, *remove @*, *remove #*), dan *filter stopwords* (*directory*).

a. *Transform Case*

Operator yang digunakan pada tahapan ini adalah untuk menyamakan ukuran huruf yang agar menjadi ukuran yang sama. Hal ini dilakukan agar terdapat keseragaman huruf pada model klasifikasi dan agar meminimalkan kesalahan dalam proses *tokenize*. Untuk hasil dari *transform case* dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1. Hasil Dari *Transform Case*

Text	<i>Transform Case</i>
Yang mau kirim PAP tit*t	yang mau kirim pap tit*t
tolong intro dulu ya.. kan	tolong intro dulu ya.. kan
jadi kaget pas buka DM	jadi kaget pas buka dm
kalo isinya barang	kalo isinya barang
ngac*ng SEMua	ngac*ng semua
#gayindonesia	#gayindonesia
#gaybandung	#gaybandung
#gbpromote	#gbpromote

b. *Tokenize*

Operator *Tokenize* untuk mengumpulkan kata dan menghilangkan tanda baca, angka, simbol, *link html*, karakter khusus atau apapun yang bukan huruf. Tabel 2 memperlihatkan perbedaan text sebelum dan sesudah proses ini.

Tabel 2. Hasil Dari *Tokenize*

Text	<i>Tokenize</i>
cari brondong top yg sang* nih daerah mtp bjb bjm ada? nanti aku bayar :) aku orangnya ganteng koq gak bakal nyesal:D #brondongsange #brondong #brondonggay #brondongcoli #gaykasel #gaybjm	cari brondong top yg sang* nih daerah mtp bjb bjm ada nanti aku bayar aku orangnya ganteng koq gak bakal nyesal

Berdasarkan Tabel 2 diatas dapat diketahui bahwa *tokenize* menghilangkan tanda baca yang tidak digunakan dan menghilangkan kata yang berawalan “#” (*hashtag*).

c. *Filter Stopwords (Directory)*

*Filter Stopwords (Directory)* yang berfungsi untuk menghilangkan kata-kata yang tidak mempunyai nilai atau kata-kata yang tidak mempunyai hubungan dengan isi tweet. Seperti kata ini, dan, atau, dan lain-lain.

Tabel 3. Hasil Dari *Filter Stopwords (Directory)*

Text	<i>Filter Stopwords</i>
jual vgak murah harga dijamin murah dan vidio banyak chat wa atau telegram	jual vgak murah harga dijamin murah vidio banyak chat wa telegram

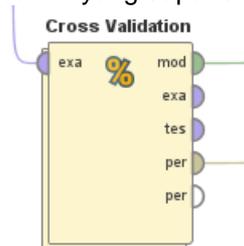
Berdasarkan Tabel 3 diatas dapat diketahui bahwa *filter stopwords* menghapus kata yang tidak mempunyai makna. Dapat dilihat kata “dan” dan “atau” karena tidak mempunyai makna sehingga kata “dan” dan “atau” dihilangkan.

3. SMOTE

Metode *Synthetic Minority Over-sampling Technique* (SMOTE) merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan cara sampling ulang sampel kelas minoritas

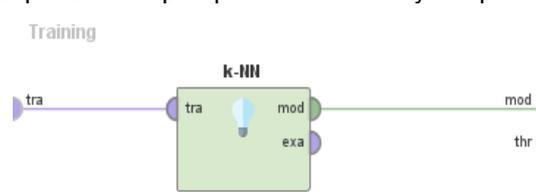
4. Proses Validasi Metode k-NN

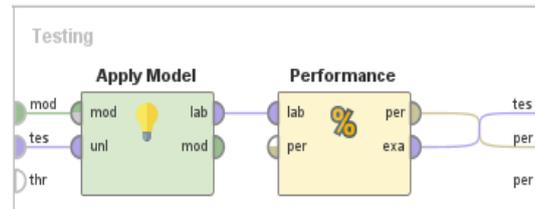
Pada proses ini menggunakan beberapa operator, sebelumnya menggunakan operator *cross validation* dengan k-10 *fold cross validation* yang dapat dilihat pada Gambar 6 dibawah ini:



Gambar 6. Proses Validasi Metode k-NN

Didalam Gambar 6 terdapat beberapa operasi di antaranya dapat dilihat pada Gambar 7.





Gambar 7. Kumpulan Operator Didalam Cross Validation

## 5. Hasil Penelitian

Penelitian ini menggunakan metode k-NN untuk melakukan analisis sentimen, dan untuk melakukan evaluasinya menggunakan *performance vector*. Dari tahapan-tahapan yang dilakukan maka menghasilkan nilai *Accuracy* yang dapat dilihat pada Gambar 8.

Tabel 4. Table View Performance Vector

Accuracy:88,25%			
	True Positif	True Negatif	
Pred.Positif	377	45	89,34%
Pred.Negatif	57	389	87,22%
	86,87%	89,63%	

Diketahui dari Gambar 4 bahwa *cross validation* k-NN mendapatkan nilai *accuracy* sebesar 88,25%, *class precision* untuk *prediction* positif sebesar 89,34%, *class prediction* untuk *prediction* negatif sebesar 87,22%, *class recall* untuk *true* positif sebesar 86,87%, dan *class recall* untuk *true* negatif sebesar 89,63%.

Berdasarkan hasil sentimen analisis dan pengujian metode k-NN maka dapat dinyatakan bahwa penggunaan *Twitter* secara bebas membuat kaum homoseksual berani menunjukkan jati diri mereka kepada masyarakat. Pengamatan secara langsung pada sentimen analisis data *tweet* yang bersifat positif ditemukan bahwa *Twitter* dijadikan sarana untuk saling berkomunikasi yang bertujuan untuk mencari pasangan sejenis, mengajak berhubungan seks, saling bertukar foto-foto seks dan masturbasi, dan penawaran pijat plus-plus (pijat yang diakhiri hubungan seks). Sedangkan untuk sentimen negatif menurut data ditemukan tweet berupa iklan penjualan, komunitas HIV AIDS, dan lain-lain. Penyebaran konten yang berkaitan dengan pornografi termasuk pornografi homoseksual dapat dengan mudah ditemukan di *Twitter*. Penggunaan *Twitter* secara umum dan bebas membuat *Twitter* dijadikan sebagai media untuk mengakses konten-konten pornografi di Indonesia. Peningkatan tersebut akan memicu munculnya kejahatan seksual dan berbagai penyimpangan-penyimpangan lain

## E. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa penelitian ini menggunakan satu *query* yang berupa *hashtag* yang berkaitan dengan pornografi khususnya dikalangan homoseksual untuk mendapatkan data dari *Twitter*. Data yang didapatkan menunjukkan bahwa *tweet* yang bersifat positif mengandung unsur pornografi sangat tinggi dengan dengan jumlah 434 data dari 615 data yang didapat-kan, sedangkan sisanya sebesar 181 data tidak mengandung unsur pornografi akan tetapi berupa iklan penjualan, informasi HIV AIDS, dan lain-lain dengan nilai *accuracy* sebesar 88,25%. Hasil penelitian ini menghimbau kepada para pengguna *Twitter* untuk lebih bijaksana, selektif, dan kritis dalam menggunakan media sosial *Twitter*.

Untuk selanjutnya bisa dijadikan bahan penelitian lain agar tidak terfokus pada satu metode saja dengan menggunakan dataset yang lebih banyak lagi sehingga dapat menghasilkan hasil yang maksimal dan akurat.

## REFERENSI

- Aputra, S. A., Didi Rosiyadi, Windu Gata, & Syepri Maulana Husain. (2019). Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 377–382. <https://doi.org/10.29207/resti.v3i3.1118>
- Anggraini, N., & Suroyo, H. (2019). Comparison of Sentiment Analysis against Digital Payment

- “T-cash and Go-pay” in Social Media Using Orange Data Mining. *Journal of Information Systems and Informatics*, 1(2), 152–163. <https://doi.org/10.33557/journalisi.v1i2.21>
- C, A. R., & Lukito, Y. (2017). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. *Jurnal ULTIMATICS*, 9(1), 50–58. <https://doi.org/10.31937/ti.v9i1.564>
- Hidayat, andi nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistem Informasi Dan Komputer (Jesik)*, 1(1), 1–7.
- Kurniawan, S., Gata, W., Puspitawati, D. A., -, N., Tabrani, M., & Novel, K. (2019). Perbandingan Metode Klasifikasi Analisis Sentimen Tokoh Politik Pada Komentar Media Berita Online. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 176–183. <https://doi.org/10.29207/resti.v3i2.935>
- News.detik.com, “Kesal Diajak Hubungan Sesama Jenis, Pria di Jaktim Bunuh Sales HP” detik newasl, 2020. [Online] Available: [https://news.detik.com/berita/d-4948400/kesal-diajak-hubungan-sesama-jenis-pria-di-jaktim-bunuh-sales-hp?\\_ga=2.160816189.1476478709.1587386483-721508057.1587386483](https://news.detik.com/berita/d-4948400/kesal-diajak-hubungan-sesama-jenis-pria-di-jaktim-bunuh-sales-hp?_ga=2.160816189.1476478709.1587386483-721508057.1587386483). [Accessed: 19-April-2020].
- News.detik.com, “Gay Ini Jual Diri dengan Tawarkan Pijat Khusus Pria” news detik, 2019. [Online] Available: [https://news.detik.com/berita-jawa-timur/d-4976675/gay-ini-jual-diri-dengan-tawarkan-pijat-khusus-pria?\\_ga=2.160816189.1476478709.1587386483-721508057.1587386483](https://news.detik.com/berita-jawa-timur/d-4976675/gay-ini-jual-diri-dengan-tawarkan-pijat-khusus-pria?_ga=2.160816189.1476478709.1587386483-721508057.1587386483). [Accessed: 19-April-2020]
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *E-Proceeding of Engineering*, 2(1), 1–7.
- Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2015). Analisis Sentimen Pada Jasa Ojek Online ... (Nugroho dkk.). 156–161.
- Pudjajana, A. M., & Manongga, D. (2018). Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter Dengan Naive Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 313–318. <https://doi.org/10.24176/simet.v9i1.1922>
- Raksanagara, R., Chrisnanto, Y. H., & Hadiana, A. I. (2016). Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naive Bayes. *Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naive Bayes*, 19–24.
- Saputra, N., Adji, T. B., & Permanasari, A. E. (2015). Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM. *Jurnal Dinamika Informatika*, 5(November), 12. <http://ojs.upy.ac.id/ojs/index.php/dinf/article/view/113>
- Saputro, P. H., Aristin, M., & Tyas, Dy. L. (2017). Berdasarkan Lirik Menggunakan Metode Tf-. *Jurnal Teknoloi Informatika Dan Terapan*, 4(1), 45–50.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor. *Journal of Intelligent Systems and Computation*, September, 43–49.
- Setiowati, Y., & Helen, A. (2018). Klasifikasi Analisis Sentimen Mengenai Hotel Di Yogyakarta. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 13(1). <https://doi.org/10.33005/scan.v13i1.1052>
- Tekno.kompas.com, “Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak” tekno kompas, 2020. [Online] Available: <https://tekno.kompas.com/read/2019/10/30/16062477/>[Accessed: 20-April-2020]
- Yulian, E. (2018). Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung. *Jurnal Matematika “MANTIK,”* 4(1), 53–58. <https://doi.org/10.15642/mantik.2018.4.1.53-58>