

## Perbandingan Algoritma *Naive Bayes Classifier* dengan *K-Nearest Neighbor (K-NN)* Pada Ulasan Aplikasi Youtube Di *PlayStore*

Azmi Ulwan Dasuki, Erizal, S.Kom., M.Kom.

Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka

Jl. Tanah Merdeka No.6, Kec. Pasar Rebo, DKI

Jakarta 13830, Indonesia.

Email: [azmi.ulwan@gmail.com](mailto:azmi.ulwan@gmail.com),

[erizal@uhamka.ac.id](mailto:erizal@uhamka.ac.id)

### Abstrak

*Youtube merupakan aplikasi yang sampai saat ini masih banyak digunakan dimana pengguna youtube ini bisa menonton dan memposting berbagai jenis video mulai dari hiburan, berita, serta wawasan secara gratis. Dengan banyaknya pengguna aplikasi Youtube, maka semakin banyak pula masyarakat yang memberikan komentar mulai dari komentar positif sampai negatif pada aplikasi Youtube di Google Play Store. Tujuan dilakukannya penelitian ini yaitu untuk menganalisis sentimen pengguna aplikasi youtube berdasarkan ulasan yang ada di google play store. Penelitian ini melewati tahap penarikan data, pelabelan data, text preprocessing, dan juga klasifikasi algoritma *Naive Bayes Classifier* dan *K-NN*. Setelah melalui tahap klasifikasi *Naive Bayes Classifier* dan *K-NN*, yaitu algoritma *K-NN* menghasilkan 81,26% dengan nilai precision sebesar 83,29% dan nilai recall sebesar 96,29% sedangkan algoritma *Naive Bayes Classifier* hanya menghasilkan nilai akurasi sebesar 78,24%, precision sebesar 82,83% dan recall sebesar 92,43%. Yang artinya dari hasil analisis sentimen tersebut, algoritma *K-NN* yang lebih unggul dibandingkan algoritma *Naive Bayes Classifier* serta pendapat masyarakat terhadap aplikasi youtube lebih banyak bersifat negatif daripada positif.*

**Kata Kunci:** *Youtube, Analisis Sentimen, Naive Bayes Classifier, K-NN*

### Abstract

*Youtube is an application that is still widely used today where Youtube users can watch and post various types of videos ranging from entertainment, news, and insights for free. With the large number of Youtube application users, more and more people are giving comments ranging from positive to negative comments on the Youtube application on the Google Play Store. The purpose of this study is to analyze the sentiment of Youtube application users based on reviews on the Google Play Store. This study went through the stages of data extraction, data labeling, text preprocessing, and also the classification of the *Naive Bayes Classifier* and *K-NN* algorithms. After going through the *Naive Bayes Classifier* and *K-NN* classification stages, namely the *K-NN* algorithm produced 81,26% with a precision value of 83,29% and a recall value of 96,29% while the *Naive Bayes Classifier* algorithm only produced an accuracy value of 78,24%, precision of 82,83% and recall of 92,43%. This means that from the results of the sentiment analysis, the *K-NN* algorithm is superior to the *Naive Bayes Classifier* algorithm and public opinion towards the YouTube application is more negative than positive.*

**Keywords:** *Youtube, Sentiment Analysis, Naive Bayes Classifier, K-NN*

## 1. PENDAHULUAN

Saat ini perkembangan teknologi banyak membawa keuntungan bagi penggunaanya,

salah satunya ialah kemudahan dalam mengakses informasi dari internet. Sehingga hal tersebut mempercepat pertumbuhan aplikasi digital seperti Instagram, Facebook,

Twitter, Youtube. Adanya aplikasi youtube disini merupakan sebuah inovasi baru dikarenakan informasi disajikan dalam format audio visual. Di era serba digital ini, penggunaan sosial media semakin meningkat. Menurut data, jumlah pemakaian sosial media sosial di Indonesia akan menjadi 127 juta pengguna pada tahun 2022 atau setara dengan 46,2% dari total populasi orang Indonesia [1] serta berdasarkan penelitian dari Sulistia Maharani Harahap bahwa penggunaan aplikasi youtube di Indonesia yaitu sebesar 88% pengguna [12].

Banyak dari pengguna youtube yang memperlumahkan fitur-fitur aplikasi, iklan yang sering muncul, bahkan eror. Sehingga untuk mempermudah proses penelitian ini, maka diperlukannya analisis sentimen dengan tujuan untuk mengetahui seberapa banyak pengguna yang berpendapat positif dan juga pendapat negatif terhadap aplikasi Youtube ini.

Pada penelitian sebelumnya yang berjudul analisis sentimen relokasi ibukota nusantara menggunakan algoritma *Naive Bayes Classifier* dan *K-NN* dengan pembobotan TF-IDF menghasilkan nilai akurasi sebesar 88,12% yang dilakukan menggunakan algoritma *K-NN* sedangkan 82,27% nilai akurasi yang didapatkan dengan menggunakan algoritma *Naive Bayes Classifiers*. Sehingga algoritma *K-NN* mempunyai nilai akurasi yang lebih baik dibandingkan algoritma *Naive Bayes Classifier* pada penelitian tersebut [4].

Seiring dengan berjalannya waktu, data mining ini selalu dikaitkan dengan perkembangan-perkembangan teknologi informasi yang bisa mengolah data dalam jumlah besar [9]. Mengingat data mining ini begitu penting, maka terdapat algoritma untuk mengolah kumpulan data yang besar dan salah satunya adalah algoritma klasifikasi. Klasifikasi itu merupakan proses dimana bisa dilakukan untuk menghubungkan masukan dengan masukan khusus dan disebut label dan kelas. Contohnya, anda dapat menilai kinerja siswa sebagai "lulus" atau "tidak lulus". Berikut merupakan algoritma-algoritma yang biasa digunakan untuk mengklasifikasikan

data yaitu, *Neural Network*, *Naive Bayes Classifier*, *K-NN*, dan juga *Support Vector Machine*.

Algoritma *Naive Bayes Classifier* sangat banyak dipakai untuk penelitian-penelitian sentimen analisis. Menurut Daifa Farah (2021), *Naive bayes classifier* itu mempunyai banyak sekali kelebihan salah satunya adalah kecepatan dalam perhitungan-perhitungan, algoritma sederhana dibandingkan algoritma lainnya serta menghasilkan nilai akurasi yang tinggi [10]. Sedangkan algoritma *K-Nearest Neighbor* termasuk algoritma yang dipakai untuk mengklasifikasi data berdasarkan data training dan *K-NN* ini juga bagian dari *supervised learning*.

Dengan demikian, penentuan algoritma sangat bergantung pada tujuan yang ingin dicapai. Dengan begitu tujuan dari penelitian ini adalah untuk mengetahui serta membandingkan dua metode analisis data yaitu *Naive Bayes Classifier* dan *K-NN* pada aplikasi youtube. Selain itu, tujuan penelitian ini juga adalah untuk membandingkan *nilai accuracy*, *precision* dan juga *recall* dari kedua metode ini pada aplikasi youtube.

## 2. LANDASAN TEORI

### 2.1 Analisis Sentimen

Analisis sentimen biasa digunakan untuk mendapatkan gambaran tentang cara menganalisis jenis ulasan, atau pandangan seseorang dari emosi yang diungkapkan berdasarkan topik tertentu. Pada dasarnya banyak bidang kehidupan sehari-hari seperti teknologi, bisnis, dan lainnya yang menggunakan analisis sentimen ini [9]. Sentimen analisis juga bisa membantu menentukan keputusan dan mengerti cara pandang atau pendapat manusia terhadap beberapa topik dan juga produk-produk tertentu. Analisis sentimen digunakan untuk mendapatkan gambaran mengenai cara menganalisis jenis ulasan, atau pandangan seseorang dari emosi yang diungkapkan sehubungan dengan topik tertentu secara otomatis. Analisis sentimen merupakan klasifikasi kelas positif dan kelas negatif dari pandangan seseorang terhadap peristiwa yang

sedang terjadi atau yang sudah terjadi. Banyak bidang dalam kehidupan sehari-hari, seperti teknologi, bisnis, dan lainnya menggunakan analisis sentimen ini [7]. Analisis sentimen tingkat kalimat (*Fined-grained Sentiment Analysis*) adalah penentuan di setiap kalimat dokumen yang kemungkinan besar mengandung kalimat yang berbeda di dalam dokumen sedangkan tingkat dokumen (*Coarse-grained Sentiment Analysis*) lebih menganggap seluruh isi dokumen sebagai sentimen positif ataupun negatif.

## 2.2 Youtube

Youtube merupakan situs media sosial sebagai hiburan virtual untuk melihat, serta membagikan video ataupun suara [13]. Saat ini aplikasi sosial media youtube sangat banyak disukai oleh masyarakat dan bisa digunakan untuk semua usia. Selain untuk menonton, youtube juga bisa mengunggah video dan juga audio yang diinginkan oleh pengguna youtube [6].

## 2.3 Text Mining

*Text mining* termasuk subjek penelitian yang disukai oleh orang banyak untuk menyelesaikan masalah serta pengambilan informasi apabila pengguna berinteraksi dengan kelompok dokumen menggunakan alat analisis yang ditemukan dalam *data mining* [12]. *Teks mining* dalam algoritma tertentu bisa membantu untuk mengklasifikasikan *text* atau dokumen kedalam kategori tertentu, contohnya kategori sentimen. Text mining termasuk proses semi otomatis untuk mencari pengetahuan dan informasi dari gabungan data besar yang berupa database dan disimpan sebagai teks sehingga *text mining* ini sangat dibutuhkan dalam kehidupan sehari-hari.

## 2.4 Text Preprocessing

*Text Preprocessing* yaitu proses mengolah data tidak tersusun jadi tersusun dengan mengubah data tersebut menjadi format yang lebih mudah atau dengan cara menghilangkan data yang tidak digunakan. Proses ini tidak boleh dilewatkan karena sangat penting dalam menganalisis sentimen.

*Text preprocessing* dilakukan melalui beberapa tahap secara berurutan, berikut adalah tahap-tahap *text preprocessing*:

### a. Cleaning

Di tahap *cleaning* dilakukan penghilangan simbol yang bukan huruf (termasuk tanda baca) yang bisa mengurangi simbol yang tidak diinginkan dan tidak relevan untuk analisis sentimen. yang perlu dihilangkan yaitu seperti angka, emoji, @, # dan juga link atau URL.

### b. Case Folding

Di tahap ini huruf yang besar secara keseluruhan diubah menjadi huruf yang kecil.

### c. Tokenizing

Di tahap *Tokenizing* digunakan untuk memisahkan kata dari suatu kalimat dengan menggunakan tanda baca dan spasi yang sudah melalui tahap *case folding* dan *cleansing*.

### d. Stopword Removal

Di tahap ini merupakan hasil *tokenizing* yaitu dilakukan pembuangan kata yang tidak memiliki pengaruh didalam proses klasifikasi sentimen dalam suatu komentar yang disebut *stop-list* seperti kata “di”, “dan”, “dengan”, “ke”, petunjuk waktu dan tanya, dsb.

## 2.5 Naive Bayes Classifier

Algoritma *Naive Bayes Classifier* adalah algoritma klasifikasi yang dapat digunakan untuk memprediksi kelas kelas yang dapat menghasilkan nilai akurasi [7]. Algoritma ini melatih data yang kemudian diproses untuk membentuk pengetahuan berupa nilai probabilitas [11].

$$P(H/X) = \frac{P(X/H) P(H)}{P(X)} \quad (1)$$

Keterangan:

$X$  = Data yang kelasnya belum diketahui

$H$  = Hipotesis data  $X$  merupakan suatu kelas spesifik

$P(H/X)$  = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$  = Probabilitas hipotesis H (prior prob.)

$P(X/H)$  = Probabilitas X berdasarkan kondisi tersebut

$P(X)$  = Probabilitas dari X

## 2.6 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* adalah teknik pengklasifikasian perkumpulan data data yang sudah diklasifikasi sebelumnya [8]. Nilai pada fitur ini tidak akan sesuai dengan perkiraan karena tingkat keakuratan menggunakan algoritma ini sangat mempengaruhi. Penerapan algoritma *K-NN* dapat menggunakan persamaan sebagai berikut:

- Mencari nilai K berupa bilangan bulat positif.
- Menghitung jarak antara data yang baru dengan data latih.

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

(2)

Keterangan:

$x$  = data sampel  
 $y$  = data pengujian  
 $D$  = jangkauan

- Mencari urutan jarak/jangkauan terdekat dengan jarak min pada nilai K.
- Label tetangga terdekat dipakai untuk memprediksi data baru, sedangkan persamaan *K-NN* dalam menghitung nilai prediksi.

$$Y = \frac{1}{K} \sum_{i=1}^K y_i$$

(3)

Diketahui:

$Y$  = Prediksi awal  
 $K$  = Total tetangga

terdekat

$Y_i$  = keluaran tetangga terdekat

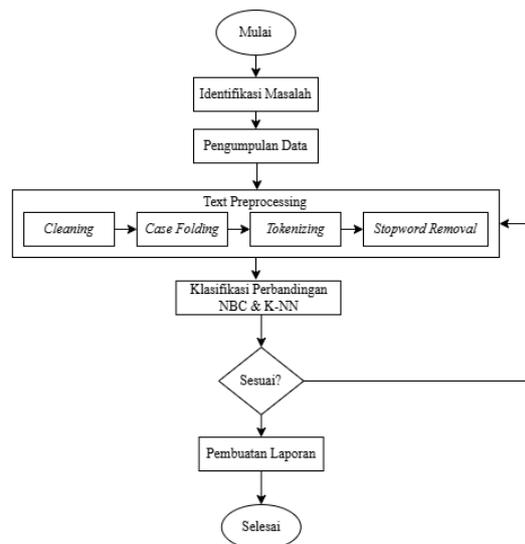
## 2.7 RapidMiner

RapidMiner merupakan perangkat lunak independen yang dirancang untuk membantu seseorang belajar ilmu data mining. Seluruh langkah-langkah yang melibatkan sejumlah data besar seperti komersial, pelatihan, bisnis, pendidikan, penelitian dan juga pembelajaran difokuskan pada *platform* ini. Perangkat lunak RapidMiner ini membantu mempelajari pengelompokan, klasifikasi dan analisis regresi serta mendukung format file seperti .xls, .csv.

## 3. METODOLOGI PENELITIAN

### 3.1 Alur Penelitian

Gambar 1 dibawah merupakan langkah-langkah yang akan dilakukan selama proses penelitian.



Gambar 1 Alur Penelitian

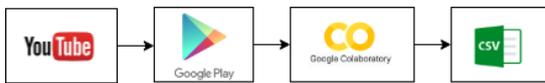
### 3.2 Identifikasi Masalah

Tahap saat ini adalah tahap awal pada metodologi penelitian untuk mengidentifikasi permasalahan mengenai ulasan dari media sosial youtube yang ada di *playstore*. Dalam penelitian ini terdapat beberapa permasalahan yang dimana harus dilakukannya penelitian berupa analisis sentimen. Permasalahan disini ialah banyak dari pengguna aplikasi media sosial youtube ini yang memberi beberapa ulasan baik itu positif dan juga negatif mengenai aplikasi youtube ini. Seperti contoh fitur-fitur yang ada di dalam aplikasinya, banyaknya iklan-iklan yang muncul,

pembayaran youtube premium yang mahal, bahkan eror dalam aplikasi.

### 3.3 Pengumpulan Data

Teknik penarikan data di penelitian saat ini yaitu menggunakan *web scrapping* yang dimana penarikan data secara otomatis dilakukan oleh kode pemrograman. Untuk mendapatkan dataset dari *playstore* membutuhkan *website Google Collaboratory* untuk ditarik datanya lalu diunduh dengan format filenya .CSV berdasarkan ketentuan data yang dibutuhkan dalam penelitian. Berikut merupakan proses pada saat melakukan teknik *web scrapping*.



Gambar 2 Proses Web Scrapping

### 3.4 Text Preprocessing

Tahapan ini menjelaskan awal proses pada teks sebelum data diolah lebih lanjut [15], berikut tahapannya:

#### a. Cleaning

Merupakan proses penghilangan karakter yang tidak berpengaruh pada tahapan analisis sentimen seperti simbol, *hashtag*, *username*, *email*, emosi dan angka .

#### b. Case Folding

Merupakan metode merubah seluruh huruf yang besar jadi huruf yang kecil.

#### c. Tokenizing

Merupakan metode pemisahan suatu kalimat menjadi kata serta menghilangkan delimiter seperti karakter pemisah dan spasi.

#### d. Stopword Removal

ialah metode pembuangan kata yang tidak berguna menggunakan *stopword*. Contohnya seperti “di”, “dan”, “dengan”, “ke”, petunjuk waktu dan tanya, dsb.

### 3.5 Klasifikasi Algoritma NBC dan K-NN

Pada tahap klasifikasi ini yaitu performa yang diukur pada metode *NBC* dan *K-NN* adalah untuk melatih data training yang sudah dilatih menggunakan operator *apply model* serta menggunakan operator *performance* untuk menampilkan nilai *accuracy*, presisi, *recall*, dan *f-measure* untuk membandingkan dan juga menarik kesimpulan berdasarkan metode mana yang lebih baik.

### 3.6 Pembuatan Laporan

Setelah dilakukannya penelitian serta mendapatkan hasil, maka akan dilakukan pembuatan laporan untuk menarik kesimpulan serta saran sebagai bahan pertimbangan.

## 4. METODOLOGI PENELITIAN

### 4.1 Identifikasi Masalah

Dari banyaknya pengguna sosial media youtube, ada beberapa masyarakat yang memberikan ulasan melalui *playstore* mengenai fitur-fitur aplikasi, banyaknya iklan yang muncul serta aplikasi eror saat digunakan. Maka dari itu ulasan tersebut harus diklasifikasikan dan dipecah menjadi dua kelas, diantaranya kelas positif dan juga kelas negatif dengan menggunakan proses klasifikasi *NBC* dan *K-NN* untuk melihat seberapa akurat dengan perbandingan dua algoritma tersebut.

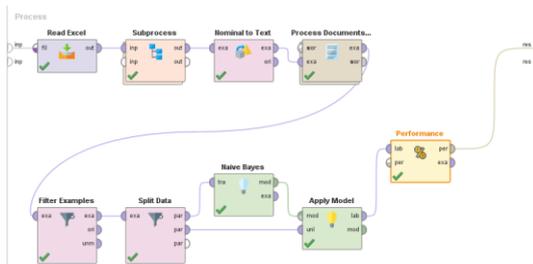
### 4.2 Pengumpulan Data

Proses pengumpulan data saat ini yaitu dengan cara menarik data memakai bahasa pemrograman *python* di *website Google Collaboratory* pada ulasan sosial media *Youtube* di *Playstore*. Data yang ditarik yaitu mulai dari tanggal 10 Mei 2024 sampai dengan 24 Oktober 2024 dan didapatkan sebanyak 1000 dataset. Gambar 3 dibawah merupakan hasil dari penarikan dataset awal.

1	userName	ulasan	label
2	Mandala P	Bug saat menaikkan resolusi sering patah patah, dan suara atau video dilay	Negatif
3	Yofi I	Update versi terbaru malah judul di video yang panjang jadi kepotong. Di klik deskripsi juga tetep kepotong	Negatif
4	Noger Perfect	Sdh di download tp tidak bisa buka.. tertulis aplikasi sudah usang .dan mencoba install kembali masih sama	Negatif
5	Athifah Silviani	Ini saya kasih bintang 2 karna 1. Sering banget iklan nya lama 2. Suka tiba² di suruh update 3. Vidio nya burem 4. Iklan nya 2-3 5. Sering ke reset sendiri nya Itu aja sih mks semoga cepat di update	Negatif
6	Rifqi Ardiansyah	Menurut saya YouTube itu aplikasi nonton yang gratis dan mudah mengakses video. Tetapi sayangnya YouTube kurang proteksi konten dewasa. Kalau bisa tolong tambahkan fitur sembunyikan konten	Negatif

Gambar 3 Hasil Penarikan data awal

Dari total keseluruhan data yang didapatkan, selanjutnya akan diberi label dan di proses secara manual oleh 3 orang dengan memanfaatkan software Microsoft Excel dan disimpan dalam format .CSV yang bertujuan untuk menganalisis respon serta pendapat masyarakat terhadap aplikasi youtube ini [14]. Setelah itu data yang sudah diberi label akan dibagi menjadi dua kelas diantaranya kelas positif dan kelas negatif. Ulasan tersebut yang sudah diberi label menghasilkan 184 komentar positif dan 810 komentar negatif. Kemudian keseluruhan data tersebut akan dibagi menjadi data training dan data testing dengan perbandingan 80:20 dengan menggunakan operator split data pada *software* rapidminer. Dapat dilihat pada gambar 4 terdapat proses model keseluruhan dalam penelitian ini.



Gambar 4 Proses Model Keseluruhan

### 4.3 Text Preprocessing

#### 4.3.1 Cleaning

Pada tahap ini dilakukannya pembersihan atau penghilangan simbol atau atribut yang tidak digunakan dalam proses *text preprocessing* seperti angka, emoji, *link* dari situs *web*, *hashtag*, @. Dapat dilihat pada gambar dibawah sebelum dan sesudah di *cleaning* [10].

label	ulasan
Negatif	Bug saat menaikkan resolusi sering patah patah, dan suara atau video dilay
Negatif	Update versi terbaru malah judul di video yang panjang jadi kepotong. Di klik deskripsi juga tetep kepotong
Negatif	Sdh di download tp tidak bisa buka.. tertulis aplikasi sudah usang .dan mencoba install kembali masih sama
Negatif	Ini saya kasih bintang 2 karna 1. Sering banget iklan nya lama 2. Suka tiba² di suruh update 3. Vidio nya burem 4. Ikl...
Negatif	Menurut saya YouTube itu aplikasi nonton yang gratis dan mudah mengakses video. Tetapi sayangnya YouTube kura...

Gambar 5 Sebelum Cleaning

label	ulasan
Negatif	Bug saat menaikkan resolusi sering patah patah dan suara atau video dilay
Negatif	Update versi terbaru malah judul di video yang panjang jadi kepotong Di klik deskripsi juga tetep kepotong
Negatif	Sdh di download tp tidak bisa bukatertulis aplikasi sudah usang dan mencoba install kembali masih sama
Negatif	Ini saya kasih bintang karna Sering banget iklan nya lama Suka tiba² di suruh update Vidio nya burem Iklan nya Seri...
Negatif	Menurut saya youtube itu aplikasi nonton yang gratis dan mudah mengakses videoTetapi sayangnya YouTube kurang ...

Gambar 6 Sesudah Cleaning

#### 4.3.2 Case Folding

Proses *case folding* ini adalah tahap pengubahan huruf yang awalnya besar diubah menjadi huruf kecil yang ada di dalam dokumen. Berikut hasil setelah dilakukannya *case folding*.

label	text
Negatif	bug saat menaikkan resolusi sering patah patah dan suara atau video dilay
Negatif	update versi terbaru malah judul di video yang panjang jadi kepotong di klik deskripsi juga tet...
Negatif	sdh di download tp tidak bisa bukatertulis aplikasi sudah usang dan mencoba install kembal...
Negatif	ini saya kasih bintang karna Sering banget iklan nya lama suka tiba² di suruh update vidio ...
Negatif	menurut saya youtube itu aplikasi nonton yang gratis dan mudah mengakses videotetapi say...

Gambar 7 Hasil Case Folding

#### 4.3.3 Tokenizing

Proses *tokenizing* ini merupakan tahapan proses pemisahan bentuk kalimat kedalam bentuk kata serta menghapus karakter yang tidak digunakan. Berikut hasil setelah dilakukan proses *tokenizing*.

label	text
Negatif	menaikkan resolusi patah patah suara video dilay
Negatif	update versi terbaru judul video panjang jadi kepotong klik deskripsi tetep kepotong
Negatif	download bukatertulis aplikasi usang mencoba install kembali
Negatif	kasih bintang karna banget iklan suka tiba suruh update vidio burem iklan reset semoga cepat update
Negatif	menurut youtube aplikasi nonton gratis mudah mengakses videotetapi sayangnya youtube kurang proteksi...

Gambar 8 Hasil Tokenizing

#### 4.3.4 Stopword Removal

Proses *stopword removal* ini adalah proses membuang kata-kata tidak berguna. Berikut hasil setelah dilakukan proses *stopword removal*.

label	text
Negatif	bug menaikkan resolusi patah patah suara video dilay
Negatif	update versi terbaru judul video panjang jadi kepotong klik deskripsi tetep kepotong
Negatif	sdh download tp bukatertulis aplikasi usang mencoba install kembali
Negatif	kasih bintang karna banget iklan nya suka tiba suruh update vidio nya burem iklan nya res
Negatif	menurut youtube aplikasi nonton gratis mudah mengakses videotetapi sayangny youtube

Gambar 9 Hasil Stopword Removal

Setelah melewati proses *text preprocessing*, dari total keseluruhan dataset sebanyak 1000 data terjadi pengurangan menjadi sebanyak 994 data dengan pembagian 184 data positif dan 810 data negatif. Kemudian total data dari hasil *text preprocessing* ini akan dibagi menjadi 795 data training dan 199 data testing.

#### 4.4 Klasifikasi Algoritma NBC & K-NN

Pada tahap ini dilakukan proses yang didalamnya terdapat operator *Naive Bayes Classifier* dan *K-Nearest Neighbor* untuk melatih data training kemudian diaplikasikan model yang sudah dilatih sebelumnya dengan menggunakan operator *Apply Model* lalu menggunakan operator *Performance* untuk memunculkan nilai akurasi. Dapat dilihat pada gambar 10 yang merupakan hasil dari akurasi yang menggunakan algoritma *Naive Bayes Classifier*.

accuracy: 78.24%

	true Negatif	true Positif	class precision
pred. Negatif	598	124	82.83%
pred. Positif	49	24	32.88%
class recall	92.43%	16.22%	

Gambar 10 Hasil akurasi NBC

Dan hasil akurasi dengan menggunakan algoritma *Naive Bayes Classifier* yaitu sebesar 78,24% dengan nilai *precision* sebesar 82,83% dan nilai *recall* sebesar 92,43%. Serta dapat dilihat hasil data dari algoritma NBC yang didapatkan dari perhitungan manual *accuracy*, *precision*, dan *recall*.

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

$$= \frac{598+24}{598+24+49+124} = \frac{622}{795} = 0,78$$

$$Precision = \frac{TP}{TP+FP}$$

$$= \frac{598}{598+49} = \frac{598}{647} = 0,92$$

$$Recall = \frac{TP}{TP+FN}$$

$$= \frac{598}{598+124} = \frac{598}{722} = 0,82$$

Dan hasil akurasi dengan menggunakan algoritma *K-Nearest Neighbor* yaitu sebesar 81,26% dengan nilai *precision* sebesar 83,29% dan nilai *recall* sebesar 96,29%. Berikut gambar hasil akurasi *K-NN*.

accuracy: 81.26%

	true Negatif	true Positif	class precision
pred. Negatif	623	125	83.29%
pred. Positif	24	23	48.94%
class recall	96.29%	15.54%	

Gambar 11 Hasil akurasi K-NN

Serta hasil data dari algoritma *K-NN* yang didapatkan dari perhitungan manual *accuracy*, *precision*, dan *recall*.

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

$$= \frac{623+23}{623+23+125+24} = \frac{646}{795} = 0,81$$

$$Precision = \frac{TP}{TP+FP}$$

$$= \frac{623}{623+125} = \frac{623}{748} = 0,83$$

$$Recall = \frac{TP}{TP+FN}$$

$$= \frac{623}{623+24} = \frac{623}{647} = 0,96$$

Sehingga hasil dari klasifikasi disini bahwa pendapat masyarakat mengenai aplikasi

youtube yang didapatkan dari ulasan yang ada di google *playstore* ini lebih dominan bersifat negatif serta algoritma *K-NN* lah yang lebih unggul dalam penelitian ini dibandingkan algoritma *Naive Bayes Classifier*.

## 5. SIMPULAN

Berdasarkan dari hasil penelitian ini maka dapat disimpulkan apabila ulasan masyarakat tentang aplikasi sosial media youtube dari *playstore* pada tanggal 10 Mei 2024 sampai dengan 24 Oktober 2024 yaitu sebanyak 1000 data kemudian terjadi pengurangan data menjadi 994 data setelah melalui tahap *text preprocessing* dengan pembagian data positif sebanyak 184 dan 810 data negatif dan juga pembagian data training dan data testing dengan perbandingan 80:20. Setelah melakukan tahap klasifikasi dengan menggunakan algoritma *Naive Bayes Classifier* maka mendapatkan nilai akurasi sebesar 78,24%, *precision* sebesar 82,83% dan *recall* sebesar 92,43%. Dan hasil dari perhitungan klasifikasi dengan menggunakan algoritma *K-Nearest Neighbor* menghasilkan tingkat akurasi sebesar 81,26% dengan nilai *precision* sebesar 83,29% dan nilai *recall* sebesar 96,29%. Sehingga ulasan masyarakat tentang aplikasi sosial media youtube lebih dominan berpendapat negatif dan algoritma *K-Nearest Neighbor* lebih akurat dalam penelitian ini dibandingkan dengan algoritma *Naive Bayes Classifier*.

## KEPUSTAKAAN

- [1] A. Roshid, S. Kom, M. Kom, R. Imada, S. Kom, dan M. Kom, "Analisis Sentimen Masyarakat Terhadap Mainan Anak 'Lato-Lato' Menggunakan Metode Support Vector Machine (Svm) Pada Media Sosial Youtube," vol. 8, no. 2502, 2023.
- [2] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, dan Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *komtekinfo*, hlm. 1–7, Jan 2023, doi: [10.35134/komtekinfo.v10i1.330](https://doi.org/10.35134/komtekinfo.v10i1.330).
- [3] A. Putri *dkk.*, "Komparasi algoritma K-NN, naive bayes dan SVM untuk prediksi kelulusan mahasiswa tingkat akhir: comparison of K-NN, naive bayes and SVM algorithms for final-year student graduation prediction," *MALCOM*, vol. 3, no. 1, hlm. 20–26, Mei 2023, doi: [10.57152/malcom.v3i1.610](https://doi.org/10.57152/malcom.v3i1.610).
- [4] D. F. Zhafira, B. Rahayudi, dan I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *JUST-SI*, vol. 2, no. 1, Agu 2021, doi: [10.25126/justsi.v2i1.24](https://doi.org/10.25126/justsi.v2i1.24).
- [5] L. Indrianingsih dan B. Budiarsih, "Analisis hukum konten negatif di platform youtube di indonesia," *HUM SOC POL*, vol. 2, no. 3, hlm. 892–916, Des 2022, doi: [10.53363/bureau.v2i3.71](https://doi.org/10.53363/bureau.v2i3.71).
- [6] R. Savira, A. Solichin, dan M. Syafrullah, "Analisis sentimen pada twitter terhadap kenaikan bbm 2022 dengan lexicon dan support vector machine," vol. 2, 2023.
- [7] R. A. Saputra, D. P. Ray, dan F. Irwiensyah, "Analisis Sentimen Aplikasi Tokocrypto Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Naïve Bayes".
- [8] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *juses*, vol. 5, no. 1, hlm. 28–37, Mei 2022, doi: [10.24246/juses.v5i1p28-37](https://doi.org/10.24246/juses.v5i1p28-37).
- [9] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k- NN untuk Analisis Sentimen Gadget Berbasis Aspek," *RESTI*, vol. 5, no. 6, hlm. 1120–1126, Des 2021, doi: [10.29207/resti.v5i6.3588](https://doi.org/10.29207/resti.v5i6.3588).
- [10] S. K. P. Loka dan A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada

- Balita: Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers,” *MALCOM*, vol. 3, no. 1, hlm. 8–14, Mei 2023, doi: [10.57152/malcom.v3i1.474](https://doi.org/10.57152/malcom.v3i1.474).
- [11] C. A. A. Soemedhy, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, dan S. Sudioanto, “Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube ‘Kekerasan Seksual’),” *JPIT*, vol. 7, no. 2, hlm. 80–84, Mei 2022, doi: [10.30591/jpit.v7i2.3547](https://doi.org/10.30591/jpit.v7i2.3547).
- [12] S. M. Harahap dan R. Kurniawan, “Analisis Sentimen Komentar Youtube Terhadap Food Vlogger Dengan Menggunakan Metode Naïve Bayes,” *MEANS*, hlm. 87–96, Jun 2024, doi: [10.54367/means.v9i1.3912](https://doi.org/10.54367/means.v9i1.3912).
- [13] A. Karimah, G. Dwilestari, dan M. Mulyawan, “Analisis sentimen komentar video mobil listrik di platform youtube dengan metode naive bayes,” *jati*, vol. 8, no. 1, hlm. 767–737, Mar 2024, doi: [10.36040/jati.v8i1.8373](https://doi.org/10.36040/jati.v8i1.8373).
- [14] H. Hidayat, F. Santoso, dan L. F. Lidimillah, “Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine),” *G-Tech*, vol. 8, no. 3, hlm. 1729–1738, Jul 2024, doi: [10.33379/gtech.v8i3.4497](https://doi.org/10.33379/gtech.v8i3.4497).
- [15] V. N. Romadina, O. Juwita, dan P. Pandunata, “Analisis Komentar Toxic Terhadap Informasi COVID-19 pada YouTube Kementerian Kesehatan Menggunakan Metode Naïve Bayes Classifier,” *INFORMAL*, vol. 9, no. 1, hlm. 92, Mei 2024, doi: [10.19184/isj.v9i1.48126](https://doi.org/10.19184/isj.v9i1.48126).