

# ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA PADA PLATFORM YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Syadam Husyen<sup>1)</sup>, Amelia Yusnita<sup>2)</sup>, Salmon<sup>3)</sup>

Program Studi Sistem Informasi, STMIK Widya Cipta Dharma  
Jl. Prof. M. Yamin No 25, Samarinda 75123  
E-mail: [2041069@wicida.ac.id](mailto:2041069@wicida.ac.id)<sup>1)</sup>, [amelia@wicida.ac.id](mailto:amelia@wicida.ac.id)<sup>2)</sup>, [salmon@wicida.ac.id](mailto:salmon@wicida.ac.id)<sup>3)</sup>

## ABSTRAK

Penelitian ini menganalisis sentimen masyarakat terhadap pemindahan Ibu Kota Negara di Platform Youtube menggunakan metode *Naïve Bayes Classifier*. Youtube, sebagai platform media sosial, menyediakan wawasan berharga tentang reaksi publik terhadap isu pemindahan ibu kota. Data komentar Youtube dikumpulkan dan dianalisis untuk mengidentifikasi sentimen positif, negatif, dan netral. Proses analisis meliputi pengumpulan data, pre-processing, pelatihan model *Naïve Bayes Classifier*, serta evaluasi menggunakan confusion matrix dan visualisasi. Hasil penelitian menunjukkan bahwa metode *Naïve Bayes Classifier* mencapai akurasi 70%, presisi 57%, dan recall 66%. Temuan ini diharapkan dapat berkontribusi pada pengembangan metode analisis sentimen yang lebih efektif dan memberikan dasar untuk penelitian lebih lanjut tentang dinamika opini publik dalam konteks politik modern.

**Kata Kunci:** Analisis Sentimen, Pemindahan, Platform Youtube, *Naïve Bayes Classifier*

## *Sentiment Analysis of the Relocation of the National Capital on the YouTube Platform Using the Naïve Bayes Classifier Method.*

## ABSTRACT

*This study analyzes public sentiment towards the relocation of the National Capital on the YouTube platform using the Naïve Bayes Classifier method. YouTube, as a social media platform, provides valuable insights into public reactions to the issue of capital relocation. YouTube comment data was collected and analyzed to identify positive, negative, and neutral sentiments. The analysis process included data collection, pre-processing, training the Naïve Bayes Classifier model, as well as evaluation using a confusion matrix and visualization. The results of the study show that the Naïve Bayes Classifier method achieved an accuracy of 70%, a precision of 66%, and a recall of 97%. These findings are expected to contribute to the development of more effective sentiment analysis methods and provide a foundation for further research on the dynamics of public opinion in the context of modern politics.*

**Keywords:** *Sentiment Analysis, Relocation, YouTube Platform, Naïve Bayes Classifier*

## 1. PENDAHULUAN

Rencana pemindahan ibu kota Indonesia menjadi topik paling hangat yang dibicarakan di berbagai kalangan. Banyak yang mengungkapkan pendapat mereka di platform media sosial, dan YouTube jadi sorotan utama. Dikenal sebagai platform populer di Indonesia, YouTube menjadi tempat para pejabat berbagi pemikiran mereka. Ini karena YouTube seringkali Dalam konteks Indonesia, YouTube menjadi tempat para pejabat berbagi pemikiran mereka. Ini karena YouTube seringkali mengangkat topik terkini, termasuk soal pemindahan ibu kota baru ini. Masyarakat yang antusias mengikuti perkembangan isu ini, menciptakan diskusi yang dinamis dan beragam di platform ini. Tidak hanya di YouTube, isu pemindahan ibu kota ini juga ramai diperbincangkan di

Twitter, tempat orang-orang secara aktif berbagi pandangan mereka. Platform ini menjadi saksi bisu dari beragam tanggapan masyarakat terhadap keputusan pemerintah. Twitter, dengan formatnya yang singkat dan langsung, memungkinkan persebaran opini dan informasi dengan cepat, memperkuat keterlibatan publik dalam isu nasional ini. Ide untuk memindahkan ibu kota negara sebenarnya bukan baru. Mulai dari zaman Presiden Soekarno hingga Susilo Bambang Yudhoyono, ide ini sudah muncul, namun belum pernah terwujud. Pada tanggal 29 April 2019, Presiden Joko Widodo membuat keputusan bersejarah untuk memindahkan ibu kota dari Pulau Jawa. Langkah ini merupakan bagian dari upaya lebih luas untuk mendistribusikan pembangunan dan pertumbuhan ekonomi secara lebih merata di seluruh

negeri Penelitian ini menarik karena menggunakan metode Naïve Bayes untuk mengklasifikasikan komentar di YouTube, terkait rencana pemindahan ibu kota. Menariknya, penelitian ini juga membandingkan tingkat akurasi dari hasil klasifikasi tersebut. Komentar-komentar di YouTube ini dikategorikan ke dalam sentimen positif atau negatif, memberikan gambaran beragam opini masyarakat. Ini mengungkapkan bagaimana teknologi dan algoritma dapat membantu memahami pandangan publik terhadap isu sosial dan politik. Di YouTube, diskusi tentang kebijakan pemindahan ibu kota ini semakin seru. Ada yang mendukung, namun tak sedikit yang menyatakan ketidaksetujuan mereka. Diversitas pandangan ini menunjukkan betapa topik ini memicu berbagai reaksi dan membuka ruang untuk dialog yang lebih inklusif dan mendalam. Text mining, atau penambangan teks, adalah teknik untuk menggali informasi dari berbagai sumber teks, seperti artikel, buku, atau komentar di media sosial. Ini merupakan bagian dari penambangan data, yang bertujuan untuk mengekstrak informasi dari data tidak terstruktur. Dalam konteks ini, text mining digunakan untuk mengidentifikasi topik utama atau gagasan dari sebuah teks. Algoritma Naive Bayes, misalnya, bisa digunakan untuk menentukan sentimen dari teks tersebut. Pemanfaatan text mining dalam studi ini membuka peluang besar dalam analisis opini publik dan pemahaman tren sosial.

## 2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

### 2.1 Cakupan permasalahan

Cakupan permasalahan berdasarkan latar belakang pada penelitian ini adalah “ Bagaimana melakukan analisis sentimen dan berapa Akurasi hasil analisis sentiment Pembangunan Ibu Kota Negara pada platform youtube menggunakan metode *Naïve Bayes Classifier*?.”

### 2.2 Batasan-batasan penelitian

Agar tidak meluasnya permasalahan maka dibuatlah batasan masalah sebagai berikut :

1. Fokus Penelitian ini pada tanggapan Masyarakat terhadap Pemindahan Ibu Kota Negara Ke Kalimantan Timur.
2. Klasifikasi sentimen dinyatakan dalam bentuk positif, negatif, dan netral.
3. Data diambil dari media sosial Youtube .

### 2.3 Rencana hasil yang didapatkan

Penelitian ini bertujuan untuk menganalisis sentimen dari komentar di Youtube dengan memanfaatkan algoritma *Naïve Bayes Classifier* dan selanjutnya untuk mengevaluasi kinerja dari *Naïve Bayes Classifier* tersebut.

## 3. BAHAN DAN METODE

Sebagai landasan dilakukannya penelitian ini, maka perlu dikembangkan teori-teori yang berhubungan dengan analisis sentimen, rencana pemindahan ibu kota, *Naive Bayes Classifier*.

### 3.1 Analisis Sentimen

Menurut Liu (Rahmawati, 2023) Analisis sentimen atau opinion mining merupakan salah satu bidang Natural Language Processing (NLP) yang menganalisis pendapat, perilaku, penilaian, dan emosi seseorang terkait dengan produk, topik, layanan, organisasi, individu, atau kegiatan lainnya yang diekspresikan ke dalam bentuk teks

### 3.2 Youtube

Menurut (Kamhar, M. Y., & Lestari, E., 2019) Media sosial, khususnya platform YouTube, telah menjadi bagian integral dari kehidupan sehari-hari masyarakat. YouTube tidak hanya sebagai hiburan, tetapi juga alat untuk memperluas pengetahuan dan menyuarakan berbagai pandangan. **Naive Bayes Classifier**

### 3.3 Naive Bayes Classifier

Menurut (Yuliana & Erlangga, 2017) Algoritma Naive Bayes mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat.

### 3.4 Metode

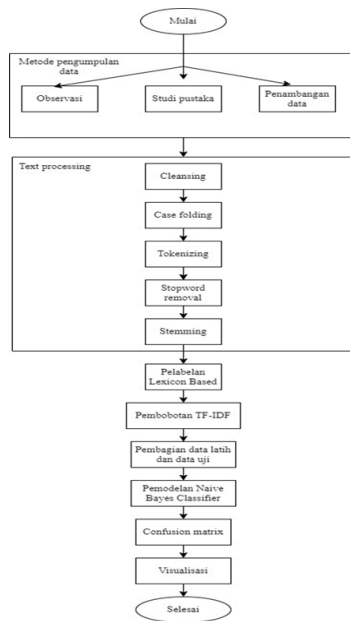
Metode penelitian merupakan sebuah cara yang digunakan untuk mendapatkan dan menganalisa data untuk kegunaan dan tujuan tertentu. Di dalam metode penelitian ini akan ada beberapa hal yang perlu diperhatikan.

### 3.1 Tempat dan Waktu Penelitian

Tempat penelitian dilakukan di BAPPEDA Badan Perencanaan Pembangunan Daerah Provinsi Kalimantan Timur Samarinda. Penelitian ini dilakukan selama kurang lebih satu bulan yang dimulai pada bulan Juni sampai Juli 2024.

### 3.2 Tahapan Penelitian

Diagram Tahapan Penelitian yang dibuat adalah rencana yang akan dilakukan oleh peneliti dalam melakukan analisis sentimen. Pada diagram tahapan penelitian terdapat rincian dalam proses pencarian awal data sampai dengan akhir, pada masing-masing proses yang ditulis pada diagram tahapan akan dijelaskan satu-persatu guna memperjelas diagram yang diolah tersebut. Bisa dilihat pada gambar. Penjelasan rinci ini mencakup semua langkah penting, mulai dari pengumpulan data mentah hingga tahap analisis dan interpretasi hasil, memastikan bahwa setiap tahapan penelitian diuraikan dengan jelas untuk memudahkan pemahaman keseluruhan proses penelitian. Langkah-langkah ini meliputi pengumpulan data, pra-pemrosesan data, metode analisis, dan lain – lain secara terstruktur dalam diagram tersebut.



Gambar 3. 1 Tahapan Penelitian

### 3.2.1 Pengumpulan Data

Penelitian ini menggunakan data kuantitatif untuk menganalisis data dari teks berbahasa Indonesia berupa komentar video yang diperoleh dari Youtube. Data serta informasi yang berkaitan dengan objek penelitian ini diperoleh melalui proses pengumpulan data sebagai berikut:

#### 1. Observasi

Observasi dilakukan sebagai studi pendahuluan, di mana peneliti mengamati secara langsung pembangunan Ibu Kota Negara dengan meninjau pendapat atau komentar yang ada pada media sosial YouTube. Observasi ini dilakukan pada bulan Juni 2024. Peneliti mengumpulkan data dari berbagai video yang membahas pembangunan, baik dari saluran resmi pemerintah maupun dari berbagai vlogger dan komentar masyarakat umum yang memberikan pandangan mereka. Beberapa hal yang diperoleh peneliti dari observasi ini adalah adanya beragam tanggapan mengenai pembangunan Ibu Kota Negara. ini memberikan wawasan awal yang berharga bagi penelitian lebih lanjut mengenai persepsi publik terhadap proyek besar ini.

#### 2. Studi Pustaka

Metode pengumpulan studi pustaka ini dilakukan untuk mencari, mempelajari, dan menggunakan berbagai literatur seperti buku, jurnal, paper, dan literatur lain yang berhubungan dengan tema penelitian ini. Informasi yang terkumpul akan dijadikan data pendukung dan pembandingan untuk penelitian ini. Literatur yang diidentifikasi akan memberikan wawasan teoretis dan kontekstual yang diperlukan untuk menganalisis sentimen terhadap pemindahan Ibu Kota Negara.

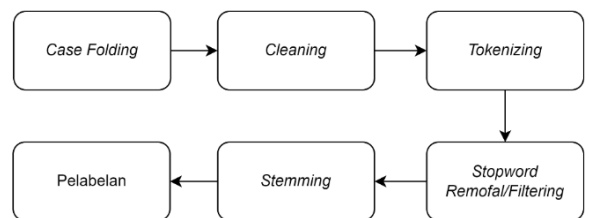
#### 3. Crawling Data

Karena sumber yang digunakan adalah YouTube, maka data dikumpulkan melalui YouTube dengan cara *crawling* data dengan memfokuskan salah satu video mengenai Pembangunan Ibu Kota Negara yang memiliki banyak komentar untuk dijadikan bahan penelitian. Penulis

melakukan *crawling* data komentar YouTube menggunakan *Export Comments* situs yang memudahkan pengguna untuk mengunduh semua komentar dari berbagai platform media sosial seperti Facebook, Twitter, Instagram, YouTube, TikTok, dan lainnya. Pengguna hanya perlu menyalin URL dari postingan publik untuk mengeksport komentar dalam format Excel atau CSV. Situs ini juga mendukung berbagai bahasa internasional dan membantu dalam analisis data komentar, memilih pemenang kompetisi dari komentar, dan memahami waktu aktivitas audiens berdasarkan waktu komentar.

### 3.2.2 Pre-Processing Data

Pre-processing merupakan tahapan awal yang akan dilalui dalam memproses teks. Pada penelitian ini akan dilakukan tahapan Pre-processing dengan tahapan case folding, data cleaning, tokenisasi, stopwords removal/filtering, stemming dan pelabelan. Alur proses pada tahapan ini ditunjukkan pada Gambar 3.2



Gambar 3. 2 Tahapan Text Processing

#### 1. Case Folding

*Case folding* adalah proses mengubah semua karakter dalam sebuah dokumen menjadi huruf kecil atau huruf besar yang sama, untuk mempercepat perbandingan selama proses pemrosesan data.

#### 2. Cleansing

*Cleansing* merupakan tahapan yang bertujuan menghilangkan karakter atau symbol link url (<http://situs.com>), username atau mention (@username), hastage(#), retweet, dan emoticon. Selain itu, pada tahap ini juga membersihkan duplikasi data.

#### 3. Tokenizing

Pada proses *tokenizing* dilakukan untuk memisahkan teks menjadi bagian-bagian yang lebih terstruktur sehingga mempermudah proses analisis selanjutnya, seperti penghapusan karakter khusus, menghitung frekuensi kata, atau membangun representasi vektor dari teks.

#### 4. Stopword Removal/Filtering

Tahap *stopword removal* membuang kata-kata tidak bermakna yang tidak mempengaruhi klasifikasi sentimen, seperti kata penghubung dan kata depan, berdasarkan kamus stopwords. Hal ini akan membantu masyarakat memahami dinamika dan keragaman pendapat yang ada di tengah-tengah mereka.

#### 5. Stemming

Tahap ini bertujuan untuk mengembalikan suatu kata ke dalam bentuk dasarnya. Dengan menghilangkan kata imbuhan awalan, akhiran, sisipan atau pun kombinasi

imbuhan. Pada penelitian ini akan digunakan library sastrawi untuk melakukan proses *stemming*. Sastrawi merupakan library pada Bahasa pemrograman python yang dibangun dengan algoritma NA. Algoritma tersebut berdasarkan aturan Bahasa Indonesia.

### 6. Pelabelan

Pelabelan adalah memberikan kategori atau tanda pada data agar atribut tertentu dapat diidentifikasi. Dengan pelabelan ini memudahkan menentukan kalimat sentimen yang terdiri dari negative positif ataupun netral. Pelabelan akan dibantu dengan kamus lexicon, berupa kamus kata positif dan negative.

### 3.2.3 Pelatihan Model Menggunakan Naive Bayes Classifier

#### 1. Pembobotan TF-IDF

Setelah tahap preprocessing berikutnya adalah tahap pembobotan kata. Pembobotan bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. TF-IDF (Term Frequency-Inverse Document Frequency) memberi nilai penting pada kata dalam satu dokumen dengan mempertimbangkan seberapa sering kata tersebut muncul di dokumen itu dan seberapa jarang kata itu muncul di seluruh kumpulan dokumen. Rumus pada TF-IDF seperti pada dibawah berikut ini :

$$TF_{(t,d)} = \frac{\text{Number of times } t \text{ appear in document } d}{\text{Total number of terms in document } d}$$

$$IDF_{(t,d)} = \log \left( \frac{\text{Total number of documents } D}{\text{Number of document the term in it}} \right)$$

$$TFIDF(t, d D) = TF(t,d) * IDF(t,d)$$

dimana :  $t_i$  = Dokumen ke - i,

$df$  = Document frequency,

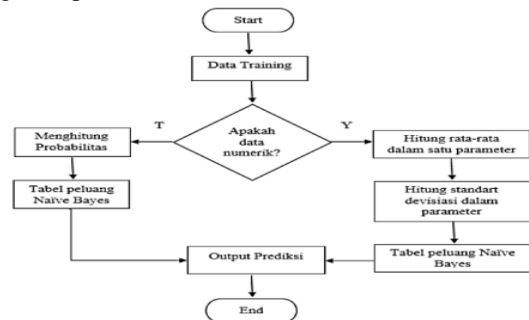
$n$  = Banyaknya data

$idf$  = Inverse document frequency.

### 3.2.4 Pemodelan Dengan Metode Naive Bayes Classifier

Metode Naive Bayes Classifier digunakan dalam penelitian ini untuk melakukan analisis sentimen, yang bertujuan untuk mengklasifikasikan teks menjadi kelas positif, negatif, dan netral berdasarkan bobot fitur dari masing-masing dokumen. Proses ini melibatkan pembobotan fitur menggunakan Naive Bayes Classifier, yang kemudian menghasilkan pola-pola yang digunakan dalam pengujian. Evaluasi dilakukan dengan memberikan skor pada setiap dokumen, di mana analisis keputusan didasarkan pada tingkat akurasi serta distribusi dokumen dalam kelas positif dan negatif. Untuk memastikan hasil yang akurat, data dibagi menjadi dua bagian, yaitu data training dan data testing, dengan menggunakan rasio 0,1. Ini berarti bahwa 90% dari data digunakan untuk proses training, di mana model Naive Bayes dilatih untuk mengenali pola dalam teks,

sementara 10% sisanya digunakan sebagai data testing untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk menghindari overfitting dan memastikan bahwa model yang dihasilkan memiliki kemampuan generalisasi yang baik pada data baru.



Gambar 3. 3 Tahap Proses Metode Naive Bayes Classifier Sumber : (Widia, 2024)

Journal of Computer and Information Systems Ampera ([journal-computing.org](http://journal-computing.org))

### 3.2.5 Confusion Matrix

Setelah hasil pengujian menggunakan Support Vector Machine selesai, akan dilakukan confusion matrix untuk menentukan 3 macam hasil performa machine learning yakni akurasi, presisi, dan recall. Hasil klasifikasi yaitu berupa jumlah sentimen positif dan negatif komentar masyarakat terhadap postingan Pemindahan Ibu Kota Negara, sehingga akan diperoleh kesimpulan mengenai tanggapan masyarakat Indonesia terhadap pemindahan Ibu Kota Negara KeKalimantan Timur.

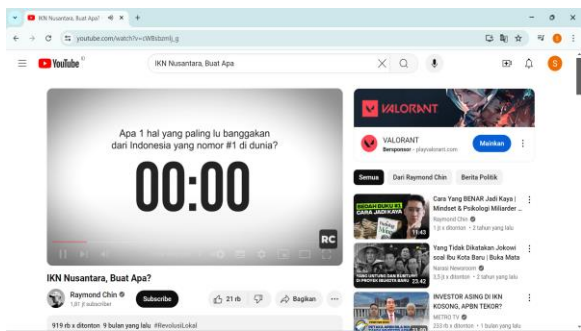
### 3.2.6 Visualisasi

Tahap ini merupakan tahap memvisualisasikan kata-kata yang sering muncul pada data komentar ke dalam bentuk wordcloud. Hal ini untuk mengetahui kata-kata apa saja yang paling banyak terdapat pada data komentar dan berapa frekuensi kata tersebut muncul dalam sebuah komentar. Visualisasi juga digunakan untuk menampilkan jumlah emosi pada data komentar dalam bentuk barplot yang terdiri atas beberapa parameter emosi.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Hasil Penelitian

Pada penelitian ini dataset yang digunakan ialah data yang berasal dari media sosial Youtube, objek penelitiannya sendiri ialah postingan dari akun Raymond Chin. Adapun tahapan proses pengolahannya hingga menjadi dataset yang nantinya akan digunakan untuk pengujian penelitian ini, berikut tahapan proses pengambilan dataset :



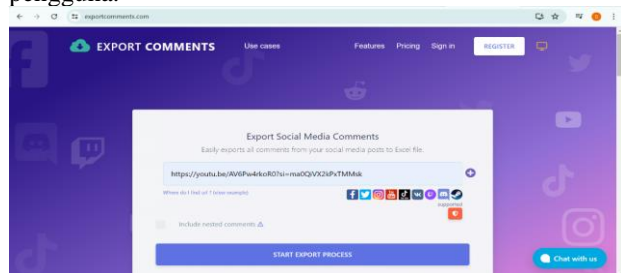
Gambar 4. 1 Youtube IKN Nusantara

## 4.2 Pembahasan

### 4.2.1 Pengumpulan Data

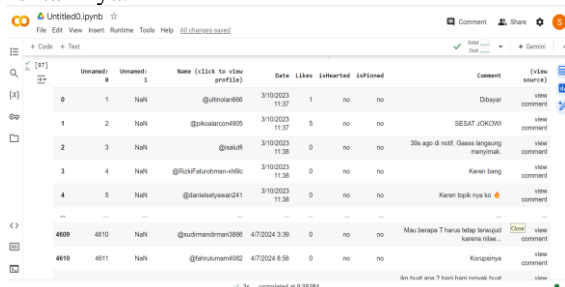
#### 1. Crawling Data

Proses *crawling* dilakukan menggunakan *Export Comment* dari Google Chrome yang secara otomatis melakukan pengumpulan data. Data yang terkumpul kemudian disimpan dalam format yang terstruktur untuk memudahkan analisis lebih lanjut. Selain itu, alat ini memastikan bahwa setiap komentar dan metadata yang relevan diambil dengan akurasi tinggi, sehingga memberikan gambaran menyeluruh tentang interaksi pengguna.



Gambar 4. 2 Proses Pengumpulan Data

Tahap ini melakukan pengambilan data komentar pada salah satu postingan pada akun Raymond Chin. Pada gambar 4.2 adalah contoh mengambil data dengan memasukkan link video youtube yang ingin kita ambil komentarnya.



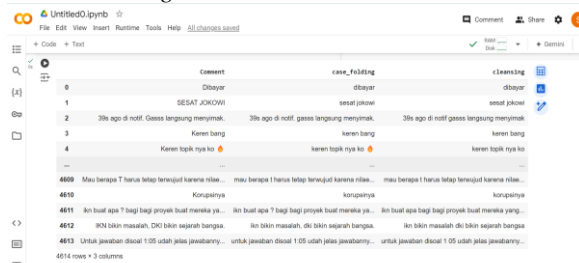
Gambar 4. 3 Hasil Crawling Data

Pada gambar 4.3 adalah hasil *crawling* data bentuk soft file csv dari jumlah keseluruhan data yang kerecord berjumlah 4614 data yang berjumlah 9 atribut diantaranya Unnamed 0, Unnamed 1, Name, Date, Like, IsHearted, IsPinned, Comment, View Source dan Date dari 9 atribut yang ditampilkan. setelah itu proses pembersihan data dan menyeleksi atribut teks.

#### 4.2.2 Pre-Processing Data

Sebelum melakukan tahapan *preprocessing* terlebih dahulu import library yang dibutuhkan untuk memanggil dataset yang akan di proses di *google collab*. Setelah library berhasil di import, selanjutnya memanggil dataset yang telah didapat dari proses *crawling* data. Pada gambar 4.3 merupakan hasil pemanggilan dataset yang berformat CSV dengan nama *comments2.csv*.

#### 1. Cleansing



Gambar 4. 4 Pembersihann Data

Pada gambar 4.4 menunjukan hasil pembersihan data pada atribut teks untuk menghilangkan atau menghapus bagian yang tidak diperlukan seperti *url*, *hashtag*, *mention*, *emoticon* dan karakter. Pada data yang sudah bersih maka akan melanjutkan ketahap *case folding*.

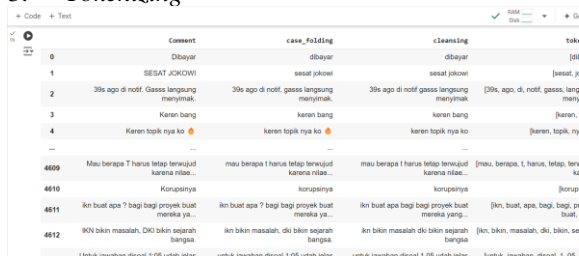
#### 2. Case Folding



Gambar 4. 5 Hasil Case Folding

Pada gambar 4.5 menunjukan hasil yang setelah dilakukan *Cleansing* dilanjut dengan *Case Folding* pada atribut teks untuk merubah semua huruf besar menjadi huruf kecil. Contoh kalimat yang ada pada gambar “SESAT JOKOWI, menjadi “sesat jokowi”.

#### 3. Tokenizing



Gambar 4. 6 Hasil Tokenizing

Pada gambar 4.6 merupakan tahapan tokenisasi setelah dilakukan *case folding* kalimat “sesat jokowi” yang jika ditokenisasikan menjadi “[‘sesat’, ‘jokowi’]” dari kalimat tersebut telah terbagi menjadi kata token yang terpisah.

#### 4. Stopword removal

Comment	case_folding	cleansing	tokenize	filtering/stopword_removal
0 Dibayar	dbayar	dbayar	[dbayar]	[dbayar]
1 SESAT_JOKOWI	sesat_jokowi	sesat_jokowi	[sesat_jokowi]	[sesat_jokowi]
2 39s ago di notif Gass langsung menyimak	39s ago di notif gass langsung menyimak	39s ago di notif gass langsung menyimak	[39s_ago, di_notif_gass_langsung_menyimak]	[39s_ago_notif_gass_langsung_menyimak]
3 Keren bang	keren bang	keren bang	[keren_bang]	[keren_bang]
4 Keren topik nya ko	keren topik nya ko	keren topik nya ko	[keren_topik_nya_ko]	[keren_topik_nya_ko]

Gambar 4. 7 Hasil Stopword removal

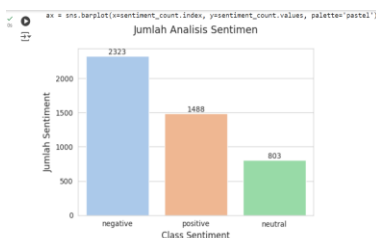
Pada gambar 4.7 merupakan tahapan stopwords removal menghilangkan kata yang tidak berguna seperti dan, pada, sebagai dll. Seperti hasil dari kalimat berikut “39s ago di notif, Gass langsung menyimak” yang jika ditokenisasikan menjadi “[‘39s’, ‘ago’, ‘di’, ‘notif’, ‘gas’, ‘la ngsung’, ‘menyimak’]” kata yang dihilangkan yaitu “[‘di’]” karena dianggap tidak terlalu memiliki arti penting dalam proses text mining dan kalimat tersebut telah terbagi menjadi kata token yang terpisah. Stemming

Comment	case_folding	cleansing	tokenize	filtering/stopword_removal	stemming_data
0 Dibayar	dbayar	dbayar	[dbayar]	[dbayar]	bayar
1 SESAT_JOKOWI	sesat_jokowi	sesat_jokowi	[sesat_jokowi]	[sesat_jokowi]	sesat_jokowi
2 39s ago di notif Gass langsung menyimak	39s ago di notif gass langsung menyimak	39s ago di notif gass langsung menyimak	[39s_ago, di_notif_gass_langsung_menyimak]	[39s_ago_notif_gass_langsung_menyimak]	39s_ago_notif_gass_langsung_simak
3 Keren bang	keren bang	keren bang	[keren_bang]	[keren_bang]	keren_bang
4 Keren topik nya ko	keren topik nya ko	keren topik nya ko	[keren_topik_nya_ko]	[keren_topik_nya_ko]	keren_topik_nya_ko
5 Cair cair	cair cair	cair cair	[cair_cair]	[cair_cair]	cair_cair
6 Pertama, saya sangat suka dengan video bang...	pertama_saya_sangat_suka_dengan_video_bang...	pertama_saya_sangat_suka_dengan_video_bang...	[pertama_saya_sangat_suka_dengan_video_bang]	[pertama_saya_sangat_suka_dengan_video_bang]	suka_video_bang_video_mu_milik_makna_nya_sahn...

Gambar 4. 8 Hasil Tahapan Stemming

Pada gambar 4.8 diatas menunjukkan proses hasil stemming dengan kata imbuhan menjadi kata dasar pada dataset, “39s ago di notif gas langsung menyimak” kata menyimak berubah menjadi kata dasar simak. Pelabelan

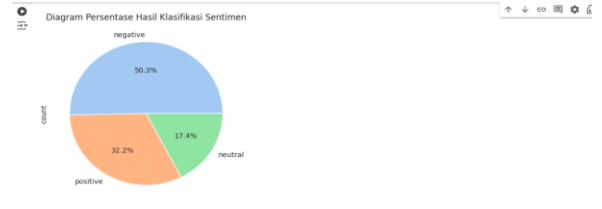
Setelah Pelaksanaan Tahapan Preprocessing dilaksanakan dan menghasilkan preprocessing dengan variable Stemming\_data yang merupakan data tidak terstruktur dan tidak memiliki kelas/label, agar dapat digunakan untuk proses mining dengan algoritma Naive Bayes Classifier, maka perlu dilakukan proses pelebelan data untuk menentukan kata-kata yang merepresentasikan kata positif dan negatif dan Menghitung jumlah kata positif dan negatif pada setiap satu kalimat dengan bantuan dari kamus inset lexicon positif dan negatif yang bersumber dari penelitian dan repository. Pelabelan diklasifikasikan kedalam 3 kelas yaitu positif, negatif dan netral.



Gambar 4. 9 Hasil Jumlah Sentimen

Pada gambar 4.9 di atas, ditunjukkan bahwa jumlah sentimen negatif lebih banyak, dengan total 2323 dokumen. Sentimen positif berjumlah 1488 dokumen, sementara sentimen netral berjumlah 803 dokumen. Hal

ini menunjukkan bahwa sentimen negatif lebih dominan dibandingkan dengan sentimen lainnya. Jika dipresentasikan, hasilnya adalah 50.3% negatif, 32.2% positif, dan 17.4% netral, seperti yang ditampilkan pada gambar 4.10. Dominasi sentimen negatif ini mengindikasikan bahwa sebagian besar teks yang dianalisis cenderung memiliki pandangan atau opini yang negatif.



Gambar 4. 10 Diagram Hasil Klasifikasi Sentimen Dalam Bentuk Persen

#### 4.2.3 Pelatihan Model Menggunakan SVM

##### 1. Pembobotan dengan TF-IDF

Setelah pelaksanaan proses pelabelan dilanjutkan dengan Tahapan Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) yang menggunakan perhitungan modul scikit-learn yaitu TfidfVectorizer. Dimana hasil dari tahap TF-IDF yaitu dari 4614 dataset setelah dilakukan pembobotan kata atau kemunculan kata berjumlah 9745 kata.

```
{'bayar': 1034,
'sesat': 8123,
'jokowi': 3912,
'39s': 194,
'ago': 396,
'notif': 6284,
'gass': 2859,
'langsung': 4573,
'simak': 8217,
'keren': 4320,
'bang': 907,
'topik': 9140,
'nya': 6329,
'ko': 4889,
'cair': 1674,
'suka': 8571,
'video': 9478,
'bang': 914,
'mu': 5819,
'milik': 5619,
'makna': 5194,
'sihh': 8202,
```

Gambar 4. 11 kamus (dictionary) yang menghubungkan kata-kata (terms) dengan indeks uniknya

Pada gambar 4.11 menjelaskan kamus ini digunakan dalam CountVectorizer atau TfidfVectorizer dari library scikit-learn, yang mengubah koleksi dokumen teks menjadi matriks istilah-dokumen. Setiap kata atau term dalam teks diberikan indeks unik untuk mengidentifikasi kata tersebut di dalam vektor atau matriks. Misalnya, kata "gass" memiliki indeks 2859, "jokowi" memiliki indeks 3912, dan seterusnya.

```
[46] #contoh memanggil salah satu kata pada kolom 1
stemming_data_tf = stemming_data[7]
print(stemming_data_tf)

['masyarakat', 'ramah', 'bang']

#hasil tf dalam dokumen 1
term_frequency = cv.transform([stemming_data_tf])
print(term_frequency)

(0, 907) 1
(0, 5359) 1
(0, 7395) 1
```

Gambar 4. 12 Hasil TF

Pada gambar 4.12 merupakan hasil TF dari kalimat “masyarakat ramah bang”, terdapat ada 3 kata dengan masing-masing nomor indeksnya. Contoh jika dalam hasil yang diberikan (0, 907) 1, ini berarti dalam dokumen pertama (indeks 907), kata “bang” muncul sekali dan begitupun seterusnya.

Word	Frequency
masyarakat	195
ramah	18
bang	2586

Gambar 4. 13 Hasil Frequency Kata

Gambar 4.13 di atas merupakan gambar “Term Frequency” yang menunjukkan frekuensi kemuculan kata-kata tertentu dalam sebuah korpus atau dataset teks

D1	TF	Idf	Tf-Idf
masyar akat	0,3333 3333	1,37404297882 2557	0,458014321 694042
ramah	0,3333 3333	2,40880508508 5081769	0,80293502
bang	0,3333 3333	0,25144906964 0699	0,08381636

Gambar 4. 14 Hasil TF-IDF

Pada gambar 4.14 menunjukkan hasil nilai yang diperoleh dari perhitungan tf-idf yang menggunakan tools google collab. Kata “masyarakat” memiliki nilai tf-idf 0,458014321694042 dan seterusnya menunjukkan pasangan indeks term dan nilai TF-IDFnya yang hasilnya dikalikan. Nilai TF-IDF memberikan bobot pada setiap term berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang term tersebut muncul di dokumen lain.

#### 4.2.4 Pemodelan Dengan Metode Naive Bayes Classifier

Setelah melewati pembobotan TF-IDF, selanjutnya tahapan pembagian data dilakukan untuk mengevaluasi performa model dari Naive Bayes, yang dibagi menjadi data *training* dan data *testing* dimana didalam penelitian ini menggunakan pembagian data 90:10. Hasil dari Nilai X train : (4152, 9627), Nilai X Test : ( 462,9627) Nilai Y train (4152) dan nilai dari Y Test (462).

```

Akurasi Linear: 70.77922077922078 %
Confusion Matrix Linear:
[[215  6 12]
 [ 41 22 17]
 [ 56  3 90]]
Precision Linear: 0.7183608348219354
Recall Linear: 0.6805912089178213
Cross Validation Linear: 0.658963739576794

[154] print(classification_report(y_test, y_pred)) # Use y_pred instead of y_pred_naive_bayes_linear

```

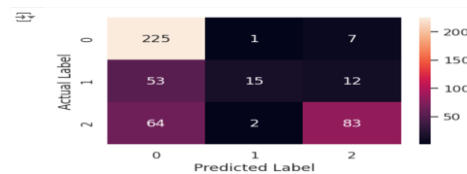
	precision	recall	f1-score	support
negative	0.69	0.92	0.79	233
neutral	0.71	0.26	0.40	89
positive	0.76	0.69	0.72	149
accuracy	0.72	0.60	0.71	462
macro avg	0.71	0.62	0.68	462
weighted avg	0.71	0.71	0.68	462

Gambar 4. 15 Hasil dari pemodelan algoritma svm

Pada gambar 4.15 diatas mengindikasikan bahwa nilai akurasi dengan pembagian 90% data training dan 10% data testing adalah 70%. Pada kelas "negative" memiliki *precision* 0.66, *recall* 0.97, dan *F1-score* 0.78. Kelas "neutral" memiliki *precision* 0.83, *recall* 0.19, dan *F1-score* 0.31, sedangkan Kelas "positive" memiliki *precision* 0.81, *recall* 0.56, dan *F1-score* 0.66.

#### 4.2.5 Confusion Matrix

Setelah didapatkannya hasil klasifikasi sentimen menggunakan Naive Bayes Classifier perlu dilakukan confusion matrix. Confusion matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi. Ini adalah alat evaluasi yang membantu mengukur kinerja model dengan menampilkan jumlah prediksi yang benar dan salah. Ini membandingkan prediksi model dengan label sebenarnya untuk memberikan gambaran mendetail tentang seberapa baik model tersebut bekerja. Perhitungan dari confusion matrix adalah accuracy, recall, dan precision. Berikut merupakan hasil evaluasi menggunakan Confusion Matrix.



Gambar 4. 16 Pengujian Confusion Matrix

Dari Gambar 4.16 diatas dapat dijelaskan bahwa memiliki 225 data sebagai negatif, 15 data sebagai data netral dan 83 data sebagai positif. Pada perhitungan manual akurasi, presisi dari matriks diatas sebagai berikut:

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar}(TP + TN + TN)}{\text{Jumlah Prediksi}(TN + FN + FP + TN)} \times 100\%$$

$$\text{Accuracy} = \frac{83 + 15 + 225}{225 + 1 + 7 + 53 + 15 + 12 + 64 + 2 + 83} \times 100\%$$

$$\text{Accuracy} = \frac{323}{462} \times 100\%$$

$$\text{Accuracy} = 0.69 \times 100\% = 70\%$$

#### 4.2.6 Visualisasi



Gambar 4. 17 Wordcloud ulasan negative

Pada gambar 4.17 merupakan hasil visualisasi dari ulasan negative Terlihat jelas bahwa frekuensi kemunculan sebuah kata meningkat seiring dengan ukurannya di dalam *word cloud*. Beberapa istilah yang sering muncul dalam ulasan Negative terhadap aplikasi yaitu “Korupsi”, “hutang”, “asing”, “bayar”, dan lainnya.



Gambar 4. 18 *Wordcloud* ulasan positive

Gambar 4.18 menampilkan hasil visualisasi ulasan positive. Dimana kata "IKN", "Indonesia", "Negara", "Jokowi", "bangga", sering digunakan dalam ulasan positive.



Gambar 4. 19 *Wordcloud* ulasan netral

Sementara itu, Gambar 4.19 menampilkan hasil visualisasi ulasan netral. Kata "rakyat", "negara", "bangun", "nan", dan "indonesia" sering digunakan dalam ulasan netral.

Hasil dari penelitian ini menggunakan data sebanyak 4614 untuk dilakukan *pre-processing*. Kemudian dilakukan pelabelan dengan hasil negative 2323 data, positive 1488 data, dan neutral 803. Kemudian dilakukan TF-IDF yang memiliki 9745 kata unik, dilanjutkan melakukan pemodelan Naive Bayes Classifier yang memperoleh hasil accuracy 70% dengan presisi dan recall untuk sentimen negatif masing-masing 66% dan 97%, untuk sentimen netral 83% dan 19%, serta untuk sentimen positif 81% dan 56%. Ini menunjukkan bahwa model Naive Bayes Classifier yang dikembangkan mampu mengklasifikasikan sentimen dengan akurasi yang cukup baik.

## 5. KESIMPULAN

Dalam penelitian ini, menentukan sentimen masyarakat dengan cara mengambil data dari ulasan atau komentar Youtube masyarakat terkait rencana pemindahan IKN Nusantara ke Kalimantan Timur kemudian data itu di

olah atau dianalisis menggunakan algoritma Naive Bayes Classifier. Sebanyak 4614 data komentar diperoleh melalui *web crawling* dari <https://exportcomments.com/>. Setelah melalui proses *Preprocessing*, di dapatkan 4614 data yang digunakan dan dibagi menjadi tiga kelas: 2323 data Negative, 1488 data Positive, dan 803 data Neutral. Penggunaan algoritma Naive Bayes Classifier pada tahap data mining mampu mengklasifikasikan komentar dengan baik. Pengujian dilakukan dengan pembagian 90:10 data *training* dan data *testing* menggunakan *confusion matrix* untuk menilai skor. Hasil terbaik diperoleh pada skenario 90:10 dengan akurasi tertinggi sebesar 70% serta nilai presisi, recal, dan F1-score yang signifikan.

## 6. SARAN

Saran untuk pengembangan lebih lanjut dari penelitian ini sebagai berikut :

1. Pada penelitian ini, *emoticon*, bahasa asing, bahasa daerah diabaikan, pada penelitian selanjutnya dapat diharapkan dapat memasukkan *emoticon*, bahasa asing dan bahasa daerah ke dalam perhitungan agar menghasilkan nilai yang lebih akurat.
2. Dapat dipertimbangkan untuk menggunakan sumber data yang lebih beragam, termasuk komentar dari media sosial lainnya.
3. Pada penelitian selanjutnya dapat menggunakan metode klasifikasi lainnya yang belum digunakan pada penelitian ini.

## 7. DAFTAR PUSTAKA

- Alifta, R. F. (2022). *Analisis Dan Klasifikasi Terhadap Pembelajaran Daring Selama Pandemi Covid 19 Berdasarkan Opini Pengguna Media Sosial Instagram Di Indonesia* (Doctoral dissertation, Universitas Komputer Indonesia).
- Herdhianto, A.(2020). *Sentiment Analysis Menggunakan Naive Bayes Classifier (Nbc) Pada Tweet Tentang Zakat*. Universitas Islam Negeri Syarif Hidayatullah.
- Kamhar, M. Y., & Lestari, E. (2019). Pemanfaat sosial media youtube sebagai media pembelajaran bahasa Indonesia di perguruan tinggi. *Inteligensi: Jurnal Ilmu Pendidikan, 1*(2), 1-7.