

Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes

Nur Hadi , Dedy Sugiarto

Program Studi Magister Teknik Industri, Fakultas Teknologi Industri, Universitas Trisakti, Jakarta, Jl. Kyai Tapa No. 1, Grogol, Jakarta Barat, 11440, Indonesia

Info Artikel

Riwayat Artikel:

Received 2024-07-08

Revised 2024-12-22

Accepted 2024-12-23

Abstract – Social media X or formerly more familiar with Twitter is one of the familiar social media and has many users in the world whis is a platform for accesing some information and commenting both suggestions and criticism related to the development of the Capital City of the Archipelago (IKN) which is the center of smart government in East Kalimantan. There are indormation, suggestions and criticisms addressed to the @ikn_id account directly addressed to the Indonesia government as well as public opinions related to IKN by using the IKN hashtag. Public sentiment on the issue is in the form of text on IKN Development. This research aims to analyze public opinion on the government's decision to build the Capital City of Nusantara (IKN) conveyed through X social media using appropriate data analysis methods by comparing the performance of support vector machine, logistic regression, and naïve bayes algorithms and identifying the most effective algorithm in sentiment analysis. The method used in this research to analyze sentiment are support vector machine, logistic regression and naïve bayes. The use of these three algorithms is also to compare the accuracy that is better than other algorithms. The results obtained using the Support Vector Machine algorithm is 80% while using the Logistic Regression and naïve bayes algorithms are 79%.

Keywords: Sentiment Analysis; Ibu Kota Nusantara; Logistic Regression; Naïve Bayes; Support Vector Machine.

Corresponding Author:

Nur Hadi

Email:

163012210008@std.trisakti.ac.id



This is an open access
article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)
license.

Abstrak – Media sosial X atau dulunya lebih familiar dengan Twitter merupakan salah satu media sosial yang familiar dan memiliki banyak pengguna di dunia yang merupakan platform untuk mengakses beberapa informasi maupun berkomentar baik itu saran maupun kritik yang berkaitan dengan pembangunan Ibu Kota Nusantara (IKN) yang merupakan pusat pemerintahan cerdas di Kalimantan Timur. Terdapat informasi, saran dan kritik yang ditujukan ke akun @ikn_id langsung yang ditujukan ke pemerintah Indonesia maupun pendapat masyarakat terkait IKN dengan memakai tagar IKN. Sentimen Masyarakat terhadap isu tersebut berupa teks terhadap pembangunan IKN. Penelitian ini memiliki tujuan untuk menganalisis opini masyarakat terhadap keputusan pemerintah dalam membangun Ibu Kota Nusantara (IKN) yang disampaikan melalui media sosial X dengan menggunakan metode analisis data yang tepat dengan cara membandingkan performa algoritma support vector machine, logistic regression, dan naïve bayes dan mengidentifikasi algoritma yang paling efektif dala analisis sentiment ini. Metode yang digunakan dalam penelitian ini untuk menganalisis sentimen adalah support vector machine, logistic regression dan baive bayes. Penggunaan tiga algoritme ini juga untuk membandingkan keakuratan yang lebih baik dari algoritma lainnya. Hasil yang diperoleh menggunakan algoritma support vector machine adalah 80% Adapun menggunakan algoritma logistic regression dan algoritma naïve bayes adalah 79%.

Kata Kunci: Analisis Sentimen, Ibu Kota Nusantara, Logistic Regression, Naïve Bayes, Support Vector Machine.

I. PENDAHULUAN

Wacana pemindahan Ibu Kota Negara Kesatuan Republik Indonesia selalu bergulir pada setiap era pemerintahan presiden, wacana ini akhirnya terealisasi pada era pemerintahan Presiden Joko Widodo [1]. Ibu Kota Negara Republik Indonesia direncanakan akan pindah ke Pulau Kalimantan pada tahun 2024 ditandai dengan akan dilaksanakannya upacara Hari Kemerdekaan Negara Kesatuan Republik Indonesia yang berlokasi di dua kabupaten di Kalimantan Timur, yaitu Kabupaten Kutai Kartanegara dan Kabupaten Penajam Paser Utara yang dinamakan Ibu Kota Nusantara (IKN). Rencana ini didasarkan pada pertimbangan bahwa kondisi Kota Jakarta dinilai tidak lagi memadai untuk berfungsi sebagai ibu kota. Gagasan pemindahan ibu kota ini telah memunculkan beragam tanggapan, baik dari kalangan politisi maupun masyarakat umum. Pemindahan Ibu Kota Negara dari DKI Jakarta ke Kalimantan Timur didasari oleh berbagai faktor, antara lain beban kota yang berlebihan, menurunnya daya dukung dan daya tampung Pulau Jawa, serta kebutuhan pemerataan pembangunan dan pengembangan pusat ekonomi baru. Jakarta menghadapi masalah pencemaran lingkungan, kemacetan parah, kepadatan penduduk, keterbatasan cadangan air

bersih, serta sistem pengelolaan air dan sanitasi yang buruk. Selain itu, banjir, ketidakseimbangan pasokan energi, tata guna lahan yang tidak memadai, serta masalah lapangan pekerjaan dan kesenjangan sosial semakin mendesak untuk diatasi [2].

Selain itu, rencana pemindahan Ibu Kota Negara dari Pulau Jawa ke Kalimantan dilatarbelakangi oleh ketimpangan sosial, ekonomi, dan kependudukan. Saat ini, sekitar 57,4% penduduk Indonesia terkonsentrasi di Pulau Jawa, sementara wilayah lain seperti Sumatera (17,9%), Bali dan Nusa Tenggara (5,5%), Kalimantan (5,81%), Sulawesi (7,31%), serta Maluku dan Papua (2,61%) memiliki sebaran penduduk yang jauh lebih kecil. Kondisi ini mencerminkan tingginya aglomerasi pembangunan di Pulau Jawa dan keteringgalan di daerah lain. Pemindahan ibu kota bertujuan untuk mengurangi beban ekologis di Jakarta yang sudah mengalami kemacetan parah dan peningkatan polusi udara serta air. Sesuai dengan Visi Indonesia 2033, ibu kota direncanakan pindah ke Kalimantan untuk menciptakan episentrum baru yang lebih dekat dengan wilayah tertinggal dan pinggiran, mendukung pemerataan pembangunan, dan mengatasi ketimpangan yang selama ini belum terjawab oleh program seperti transmigrasi dan percepatan pembangunan daerah [3]. Selain faktor sosial, ekonomi dan lingkungan, lokasi Ibu Kota Negara (IKN) yang strategis di tengah wilayah Indonesia, didukung oleh pesatnya perkembangan teknologi informasi dan transportasi, memberikan peluang besar untuk terciptanya entitas budaya baru yang lebih dinamis. Hal ini menuntut kesiapan pemerintah untuk melestarikan tradisi budaya yang telah ada, sekaligus mengembangkan sinergi dengan budaya baru yang akan terbentuk, sehingga kekayaan budaya bangsa tetap terjaga dan mampu beradaptasi dengan perubahan zaman [4].

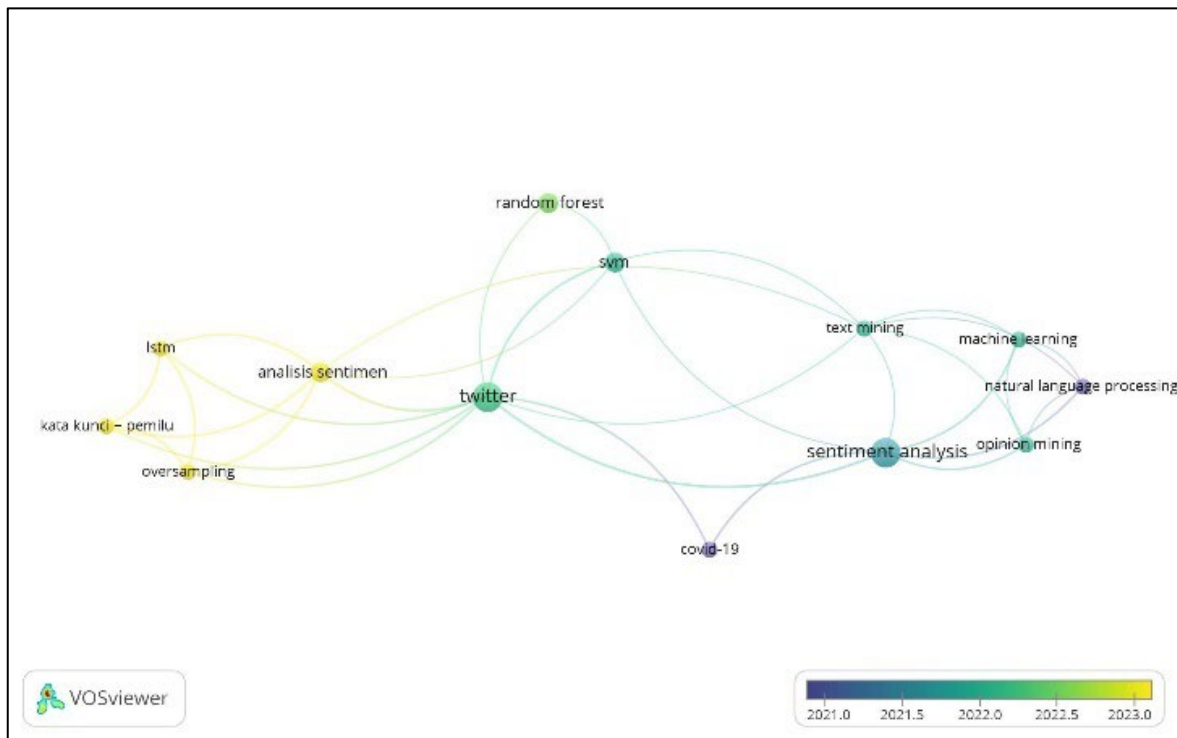
Pada tahun 2022, Negara Kesatuan Republik Indonesia menetapkan Undang-Undang No. 3 Tahun 2022 yang mengamanatkan pembangunan Ibu Kota Nusantara di Pulau Kalimantan. Langkah ini memiliki tujuan untuk memindahkan pusat pemerintahan dari Daerah Khusus Ibukota Jakarta ke Pulau Kalimantan, mencetuskan beragam reaksi dan opini dari berbagai segmen masyarakat. Dengan cakupan dan dampak yang luas, perpindahan ini tidak hanya menandai transformasi infrastruktur, tetapi juga memiliki implikasi sosial, ekonomi, dan lingkungan yang signifikan. Analisis sentimen terkait pembangunan Ibu Kota Nusantara di Pulau Kalimantan menjadi esensial untuk memahami pandangan, sikap, dan reaksi masyarakat terhadap keputusan ini yang tentunya memiliki pro dan kontra terhadap keputusan pemerintah tersebut. Melalui pendekatan ini menganalisis sentimen masyarakat dengan pendekatan *machine learning* akan melalui empat proses yaitu pengumpulan data, pra pemrosesan data, analisis dan evaluasi penelitian [5], kita bisa mengetahui informasi lebih lanjut terkait bagaimana berbagai pihak menerima atau menolak perubahan ini, serta faktor apa yang mendasari persepsi mereka. Dengan perkembangan teknologi menganalisis sentimen dapat menggunakan pendekatan *machine learning* dimana ini bisa dikategorikan dalam *natural language processing* yaitu suatu proses mengambil dan mengolah data opini masyarakat melalui berbagai media sosial salah satunya adalah media sosial X merupakan suatu hal yang menarik dalam hal mengidentifikasi dan memahami perasaan seseorang terhadap suatu permasalahan maupun fenomena dalam bentuk membuat suatu *tweet* dalam media sosial X dimana *natural language processing* sudah banyak digunakan dalam penelitian berbagai bidang [6], [7], [8], [9].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pembangunan Ibu Kota Negara (IKN) melalui data dari media sosial X dengan menggunakan metode analisis data yang tepat dengan cara membandingkan performa algoritma *support vector machine (SVM)*, *logistic regression*, dan *naïve bayes* dalam mengklasifikasikan sentimen positif, netral, dan negatif serta mengidentifikasi algoritma yang paling efektif dalam mendukung analisis sentimen Pembangunan IKN. Penelitian ini diharapkan dapat mengidentifikasi algoritma analisis sentimen yang paling efektif, sehingga dapat menjadi acuan dalam penelitian lanjutan atau aplikasi serupa. Penelitian ini juga diharapkan dapat memberikan wawasan mendalam mengenai sentimen masyarakat terhadap pembangunan Ibu Kota Negara (IKN) untuk membantu pemangku kepentingan memahami opini publik secara lebih terstruktur. Selain itu, hasil penelitian diharapkan memberikan kontribusi nyata bagi pengambilan keputusan dan penyusunan strategi komunikasi publik yang lebih responsif terhadap aspirasi masyarakat.

Dalam penelitian ini akan dilakukan beberapa langkah dari awal pengambilan data dengan cara *crawling* data *tweet* menggunakan *tweet harvest* yang menghasilkan data berupa pendapat atau perasaan yang dikemukakan oleh masyarakat, selanjutnya dilakukan *pre processing* data dengan berbagai proses dengan menghasilkan data yang sudah bagus [10], selanjutnya dilakukan proses pelabelan terhadap opini tersebut menggunakan InSet Lexicon Indonesia yang mengategorikan opini berdasarkan kategori positif, negatif ataupun netral dimana data tersebut juga akan digunakan untuk membagi data *training* dan data *testing* dan baru memodelkan data ini menggunakan algoritma *support vector machine*, *logistic regression* dan *naïve bayes* dengan nilai akurasi masing-masing dari setiap algoritma.

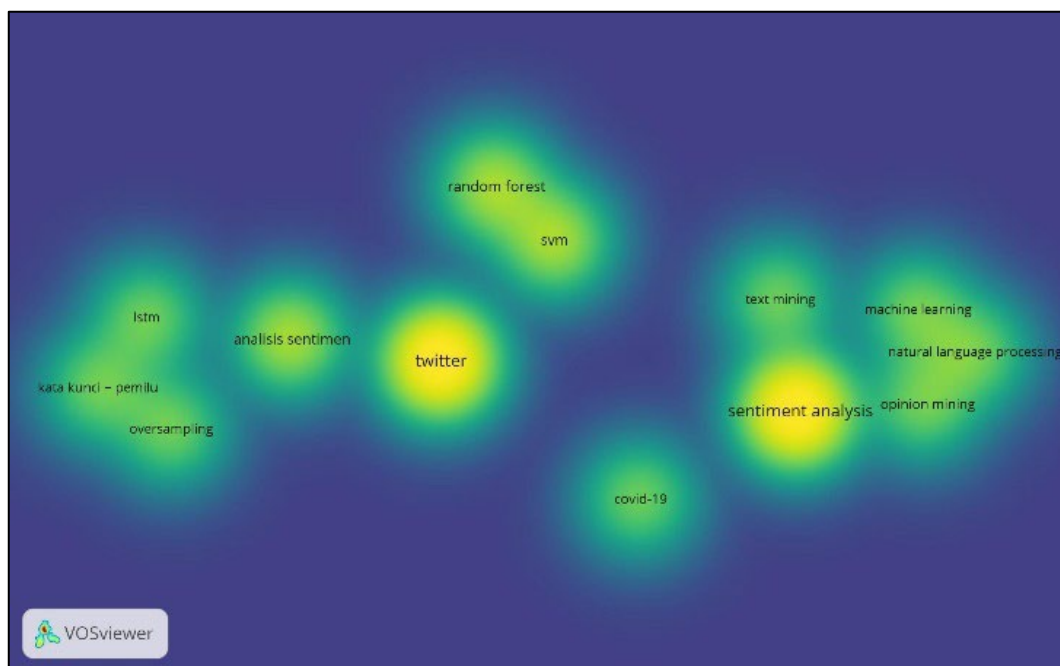
Dalam penelitian ini pada studi literatur awal dilakukan kajian terhadap beberapa artikel ilmiah terdahulu yang memiliki keterkaitan dengan penelitian yang akan diteliti yaitu terkait analisis sentimen masyarakat yang diungkapkan pada media sosial khususnya media sosial X dengan pendekatan *machine learning* yang ditampilkan sesuai hasil pengolahan jurnal melalui *Vosviewer* yang menunjukkan beberapa penelitian sebelumnya. Berdasarkan Gambar 1 *overlay visualization* menggambarkan bawah penelitian analisis sentimen dalam pendekatan *machine learning* telah dilakukan pada penelitian sebelumnya terkait perkembangan bidang ilmu pengetahuan dengan *trend* penelitian dari tahun 2021 – 2023. Penelitian terkait dapat digunakan di berbagai bidang ilmu pengetahuan maupun di berbagai isu

yang berkembang di masyarakat, dalam hal melakukan analisis sentimen ada beberapa algoritma yang pernah digunakan dalam penelitian seperti algoritma *support vector machine* maupun *random forest*.



Gambar 1. *Overlay Visualization Analisis Sentimen Berbasis Machine Learning*

Berdasarkan Gambar 2 *density visualization* pembahasan tentang analisis sentimen pada media twitter dan variabel lainnya saling terkait dan menggambarkan penelitian terdahulu. Dalam melakukan penelitian analisis sentimen terdapat beberapa metode algoritma yang dapat kita gunakan seperti pada menggunakan metode *support vector machine (SVM)*, *logistic regression* dan *naïve bayes*. Seperti penelitian yang diteliti oleh Amiruddin, dkk [6] yang menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* untuk dilakukan analisis sentimen terhadap gelombang PHK massal di Indonesia yang menyebabkan terdapatnya banyak pengangguran.



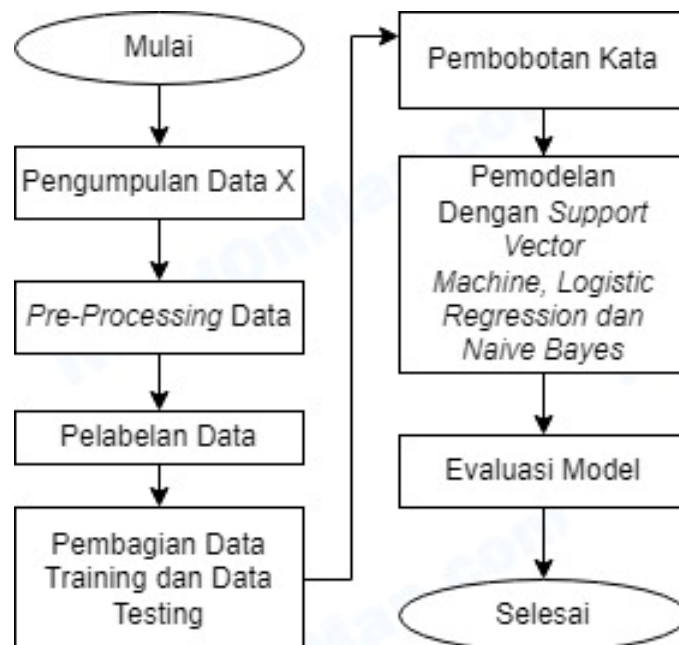
Gambar 2. *Density Visualization Analisis Sentimen Berbasis Machine Learning*

Pada penelitian Anisa, dkk [11] dalam melakukan analisis sentimen menggunakan data komentar dari youtube mengenai komentar masyarakat terhadap BTS dengan melakukan pelabelan terbagi menjadi pelabelan kategori positif maupun pelabelan yang masuk ke dalam kategori negatif menggunakan perbandingan metode *support vector machine* dan *Naive Bayes* yang menunjukkan hasil akurasi algoritma *SVM* lebih baik dibandingkan dengan algoritma *naive bayes*. Analisis sentimen yang diteliti oleh choirul, dkk [12] yang membahas analisis sentimen data twitter jelang pemilu 2024 menggunakan metode *LSTM* dalam pemodelan data menggunakan pembagian sentimen berdasarkan kategori mendukung, tidak mendukung dan netral untuk diterapkan aplikasi *deep learning* berbasis web yang akan memperlihatkan grafik dari *output* pemodelan data tersebut.

Selanjutnya hasil penelitian terhadap komparasi beberapa algoritma yang dilakukan oleh Chandra, dkk [13] yang membahas terkait kekerasan seksual pada komentar youtube dengan hasil algoritma *naive bayes* memiliki nilai akurasi sebesar 78% dan lebih baik dibandingkan algoritma *SVM*. Selain itu juga ada penelitian yang diteliti oleh ryo, dkk [14] terkait dengan migrasi TV digital yang ditetapkan oleh pemerintah melalui Kementerian Komunikasi dan Informatika Republik Indonesia dimana pada penelitian ini salah satunya juga menggunakan algoritma *logistic regression* untuk menilai sentimen masyarakat dengan hasil akurasi 90% lebih baik dari algoritma *multinomial naive bayes*. Dari beberapa penelitian terdahulu yang terkait, memiliki tiga algoritma dengan akurasi yang lebih baik dari setiap penelitian yaitu menggunakan algoritma *support vector machine*, *logistic regression* dan *naive bayes*. Maka dalam penelitian ini akan menerapkan tiga algoritma tersebut untuk menguji algoritma mana yang memiliki nilai akurasi yang baik pada analisis sentimen terhadap pembangunan IKN dengan harapan dapat mengetahui respon atau stigma masyarakat terhadap isu pembangunan IKN ini berbasis data dari media sosial X.

II. METODE

Penelitian ini menggunakan metode kuantitatif dengan mengumpulkan sejumlah data pada media sosial X dengan kata kunci "IKN" dengan melalui beberapa tahapan dimulai dari pengumpulan data di media sosial X, selanjutnya *pre-processing* data, pelabelan data, pembobotan kata, pemodelan dan evaluasi yang tahapan penelitiannya seperti pada Gambar 3.



Gambar 3. Tahapan Penelitian.

A. Data Collection

Penelitian ini menggunakan data yang diperoleh melalui metode *crawling* di media sosial X sebagai sumber data dengan memanfaatkan *Application Programming Interface (API)* yang terdapat dari akun X. Proses *crawling* dilakukan menggunakan *Tweet-Harvest* dengan tutorial yang ditulis oleh Helmi Satria pada tanggal 30 Maret 2024. Data dikumpulkan menggunakan perangkat laptop dengan kata kunci pencarian "IKN" pada tanggal 2 Juli 2024, dengan batas maksimum pengambilan sebanyak 2.000 tweet. Hasil *crawling* disimpan dalam format file .csv dan mencakup 2.008 tweet yang berisi beragam komentar masyarakat terkait pembangunan Ibu Kota Negara (IKN) [15]. Data ini mencakup teks mentah (*raw data*) yang akan diolah lebih lanjut untuk analisis sentimen, sebagaimana ditampilkan pada Gambar 4.

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name
0	1807774233231229241	Tue Jul 02 07:25:54 +0000 2024	0	@__LOVE_AG4EVER @jokowi Jadi gak nih 17an di ikn	1808039381329994096	NaN	__LOVE_AG4EVER
1	1807951726785581349	Tue Jul 02 07:25:54 +0000 2024	0	@jeffsatur_Italy Like I always proud that he w...	1808039379539001628	NaN	Hi_TheM_
2	1808039342763565349	Tue Jul 02 07:25:45 +0000 2024	0	Perkembangan tol IKN Ganta Bali Cantik Gaji 50...	1808039342763565349	https://pbs.twimg.com/ext_tw_video_thumb/18080...	NaN
3	1808039315655872865	Tue Jul 02 07:25:39 +0000 2024	0	Bandara VVIP IKN Beroperasi Fungsional 1 Agust...	1808039315655872865	NaN	NaN
4	1808039301898531100	Tue Jul 02 07:25:35 +0000 2024	0	mas @gibran_tweet ini area lahan pertanian di ...	1808039301898531100	NaN	NaN

Gambar 4. Hasil Raw Data

Setelah raw data di dapat, maka selanjutnya dilakukan *drop* data yang tidak terkait seperti data *conversation_id_str*, *created_at*, *favorite_count*, *id_str*, *image_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str* dan *username*. Selain *drop* data tersebut, juga dilakukan penghapusan *tweet* yang *double* yaitu dengan menggunakan *drop_duplicates*. Dari hasil *drop* data tersebut maka kolom yang digunakan adalah kolom '*full_text*' saja dengan jumlah *tweets* 1982 *tweets* seperti yang ditampilkan pada Gambar 3 yang akan digunakan untuk dimodelkan.

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name
0	1807774233231229241	Tue Jul 02 07:25:54 +0000 2024	0	@__LOVE_AG4EVER @jokowi Jadi gak nih 17an di ikn	1808039381329994096	NaN	__LOVE_AG4EVER
1	1807951726785581349	Tue Jul 02 07:25:54 +0000 2024	0	@jeffsatur_Italy Like I always proud that he w...	1808039379539001628	NaN	Hi_TheM_
2	1808039342763565349	Tue Jul 02 07:25:45 +0000 2024	0	Perkembangan tol IKN Ganta Bali Cantik Gaji 50...	1808039342763565349	https://pbs.twimg.com/ext_tw_video_thumb/18080...	NaN
3	1808039315655872865	Tue Jul 02 07:25:39 +0000 2024	0	Bandara VVIP IKN Beroperasi Fungsional 1 Agust...	1808039315655872865	NaN	NaN
4	1808039301898531100	Tue Jul 02 07:25:35 +0000 2024	0	mas @gibran_tweet ini area lahan pertanian di ...	1808039301898531100	NaN	NaN

Gambar 5. Output Drop Data

B. Pre-processing Data

Setelah didapatkan data *crawling* dari media sosial X dalam bentuk file .csv, selanjutnya dilakukan *pre-processing* data yaitu mengeliminasi informasi yang tidak jelas dan tidak relevan bagi penelitian ini untuk dihapus dengan berbagai langkah dalam menghasilkan data yang baik untuk dilakukan analisis [16].

C. Case Folding

Case folding merupakan proses konversi seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) atau huruf besar (*uppercase*). *Case folding* memiliki tujuan untuk menghapus perbedaan antara huruf besar dan kecil, sehingga teks bisa dianalisis ataupun diproses tanpa perlu mempertimbangkan perbedaan tersebut. Langkah ini umum dilakukan dalam pra pemrosesan teks, terutama dalam bidang *Natural Language Programming (NLP)* atau *text mining* [17].

```
Case Folding Result :
0    @__love_ag4ever @jokowi jadi gak nih 17an di ikn
1    @jeffsatur_italy like i always proud that he w...
2    perkembangan tol ikn ganta bali cantik gaji 50...
3    bandara vvip ikn beroperasi fungsional 1 agust...
4    mas @gibran_tweet ini area lahan pertanian di ...
Name: Text Case Folding, dtype: object
```

Gambar 6. Output Text Case Folding

D. Tokenizing

Text tokenizing (tokenisasi teks) adalah proses membagi teks menjadi unit-unit kata kecil yang disebut token. Token biasanya berbentuk kata-kata atau simbol tertentu yang memiliki makna penting dalam proses analisis teks.

Tokenisasi merupakan langkah fundamental dalam hal pemrosesan bahasa alami (NLP) dan diimplementasikan pada berbagai aplikasi seperti analisis teks, klasifikasi teks dan lainnya [17].

```
0      [loveagever, jadi, gak, nih, an, di, ikn]
1      [italy, like, always, proud, that, he, will, r...
2      [perkembangan, tol, ikn, ganta, bali, cantik, ...
3      [bandara, vvip, ikn, beroperasi, fungsional, a...
4      [mas, tweet, ini, area, lahan, pertanian, di, ...
Name: Text Tokenizing, dtype: object
```

Gambar 7. Output Text Tokenizing

E. Normalization

Text normalization (normalisasi teks) adalah proses mengubah teks menjadi format standar yang lebih konsisten dan seragam untuk memudahkan analisis lebih lanjut. Tujuan normalisasi teks adalah mengurangi variasi dalam teks yang tidak relevan atau dapat mengganggu analisis untuk memperoleh bentuk dasar dari kata tersebut [18].

```
0      [loveagever, jadi, enggak, nih, an, di, ikn]
1      [italy, like, always, proud, that, he, will, r...
2      [perkembangan, tol, ikn, ganta, bali, cantik, ...
3      [bandara, vvip, ikn, beroperasi, fungsional, a...
4      [mas, tweet, ini, area, lahan, pertanian, di, ...
Name: Text Normalization, dtype: object
```

Gambar 8. Output Text Normalization

F. Stemming

Text stemming merupakan suatu proses mengurangi kata-kata kedalam bentuk dasarnya atau akarnya dengan menghapus akhiran atau imbuhan yang tidak diperlukan. *Text stemming* ini memiliki tujuan untuk mengelompokkan kata yang mempunyai makna dasar sama agar kata yang akan dianalisis lebih efektif. Misalnya, kata-kata seperti "melambung" diubah menjadi kata "lambung" atau kata "menutupi" diubah ke bentuk dasar "tutup" [17]. Dalam proses *stemming* digunakan *library* sastrawi dengan *output stemming* seperti terlihat pada Gambar 9.

```
melambung : lambung
menutupi : tutup
antibiotic : antibiotic
recommend : recommend
dimension : dimension
assent : assent
```

Gambar 9. Output Text Stemming

G. Filtering

Text filtering merupakan suatu proses menghilangkan komponen-komponen yang tidak diinginkan atau tidak relevan dari teks untuk keperluan analisis. Langkah ini penting dalam pra-pemrosesan teks yang bertujuan meningkatkan kualitas data teks sebelum digunakan dalam analisis lebih lanjut atau sebagai input untuk model *machine learning*.

```
0      [loveagever, an, ikn]
1      [italy, like, always, proud, that, he, will, r...
2      [kembang, tol, ikn, ganta, bal, cantik, gaji, ...
3      [bandara, vvip, ikn, operasi, fungsional, agus...
4      [mas, tweet, area, lahan, tani, ikn, posisi, d...
Name: Text Filtering, dtype: object
```

Gambar 10. Output Text Filtering

Data dari hasil *text filtering* ini yang akan digunakan dianggap data yang sudah dilakukan *pre-processing* data yang datanya siap untuk dilakukan pelabelan dan pemodelan analisis sentiment.

H. Pelabelan Data

Setelah dilakukan *pre-processing* data, maka data yang digunakan untuk analisis adalah data *text filtering* yang sudah diolah dan dilanjutkan dengan melakukan pelabelan data menggunakan pelabelan otomatis berbasis *Inset* (*Indonesia Sentiment Lexicon*) berdasarkan referensi Fajri Koto dan Gemala Y. Rahmaningtyas pada *IEEE in the 21st International Conference on Asian Language Processing (IALP)* di Singapore pada bulan Desember tahun 2017. Pada proses pelabelan data ini juga dimasukkan kamus bahasa positif dan kamus bahasa negatif untuk mengkategorikan

suatu kata masuk ke kategori apa yang akan menghasilkan data berdasarkan kategori positif, negatif ataupun masuk kategori netral.

I. Pembagian Data Pelatihan dan Pengujian

Setelah semua data dilabelkan menggunakan *Indonesia Sentiment Lexicon (Inset)* maka data tersebut akan menghasilkan data berdasarkan kategori netral, negatif maupun positif. Selanjutnya data tersebut digunakan untuk dibagi menjadi data pelatihan dan pengujian. Data yang dilatih sebanyak 80 % dari keseluruhan data sejumlah 1585 *tweets* dan sisa data 20% sejumlah 397 *tweets* akan dilakukan untuk tes pengujian dari suatu model. Dalam membagi data pelatihan dan data uji juga diperlukan keseimbangan karena akan berdampak positif pada kinerja model pengujian [19].

J. Pembobotan Kata

Setelah didapatkan data latih, langkah selanjutnya adalah melakukan pembobotan kata, proses ini merubah kata-kata menjadi numerik, dalam penelitian ini untuk proses pembobotan kata menggunakan *Term Frequency – Inverse Document Frequency (TF-IDF)* untuk menentukan seberapa penting suatu kata dan frekuensi kemunculannya dalam dokumen [20], [21].

K. Pemodelan Data

Proses pembobotan data selesai, setelahnya dilakukan pemodelan data pengujian menggunakan tiga algoritma, yaitu algoritma *support vector machine*, *logistic regression* dan *naïve bayes*.

1) Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode pembelajaran mesin klasik yang berfokus pada klasifikasi diciptakan oleh Alexey ya. Chervonenkis dan Vladimir N. Vapnik pada tahun 1963 [22]. Metode ini berawal dari ide untuk menyelesaikan bentuk ganda dari masalah berukuran besar, sehingga pengklasifikasian hanya membutuhkan sejumlah kecil vektor pendukung untuk mencapai prinsip minimalisasi risiko struktural. Teori pembelajaran statistik digunakan untuk mengatasi masalah nonlinier dan masalah minimum lokal. Panggilan sistem dapat menggunakan frekuensi pendek untuk mengubah urutan menjadi urutan panggilan dengan panjang vektor tertentu dalam ruang berdimensi tinggi. Dengan demikian, deteksi anomali dapat dilakukan dengan menggunakan mesin vektor pendukung [23]. Dalam mengklasifikasikan menggunakan algoritma *support vector machine* pada awal penelitian ini dilakukan evaluasi terhadap 3 jenis kernel yaitu linear, polinomial dan *radial basis function*. Berikut penjabaran terkait beberapa jenis kernel tersebut [21] :

- Kernel Linear

Kernel linear menghitung hasil kali titik (*dot product*) dari dua vektor *input* dalam ruang asli tanpa mengubahnya ke ruang fitur yang memiliki dimensi lebih tinggi.

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y}) \quad (1)$$

- Kernel Polinomial

Kernel polinomial menghitung nilai dua vektor input dalam ruang asli dalam hubungan polinomial.

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + c)^d \quad (2)$$

- Kernel *Radial Basis Function (RBF)*

Kernel *Radial Basis Function (RBF)* menggunakan fungsi *gaussian* untuk menghitung diantara dua vektor input memiliki kesamaan di dalam ruang fitur.

$$K(\mathbf{x}, \mathbf{y}') = \exp(-\gamma(\mathbf{x} - \mathbf{y}'))^2 \quad (3)$$

Dengan parameter *C* adalah *cost*, *gamma* adalah parameter yang mengindikasikan kemiringan atau perubahan output terkait dengan perubahan input, *c* merupakan koefisien dan *d* merupakan derajat polinomial. *Gamma* menentukan seberapa curam perubahan *output* terhadap *input*, sedangkan koefisien menunjukkan titik potong dari suatu model. Derajat polinomial menunjukkan tingkat polinomial dari suatu fungsi tersebut [21].

2) Logistic Regression

Regresi logistik merupakan algoritma klasifikasi pada pembelajaran mesin yang digunakan dalam hal memprediksi probabilitas variabel dependen kategori. Metode ini merupakan bentuk umum dari regresi linier yang memiliki tujuan untuk melihat hubungan antara beberapa variabel bebas dengan variabel terikat yang bersifat biner atau *probabilistic* [17]. Selain itu regresi logistik juga memiliki tujuan memperoleh model yang paling sederhana dan terbaik dan dapat menjelaskan hubungan antara keluaran dari variabel respons (*y*), berikut rumus algoritma *logistic regression* [24].

$$\sum_{i=1}^m \frac{ypred^{(i)} == ytrue^{(i)}}{m} \quad (4)$$

3) Naïve Bayes

Metode klasifikasi *naïve bayes* merupakan teknik didalam analisis sentimen. Pendekatan ini memiliki keunggulan teoritis dalam hal konsistensi data dan akurasi klasifikasi. *Naïve bayes* sering digunakan dalam teknik klasifikasi, terutama di *platform* media sosial [17]. *Naïve Bayes* merupakan metode yang dikembangkan dari algoritma CART (*Classification and Regression Trees*), yang merupakan bagian dari teknik pohon keputusan. Metode ini mampu memproses data dengan cepat dan memungkinkan untuk membangun sejumlah pohon yang diinginkan. Penelitian ini akan menggunakan pengujian dengan validasi menggunakan *confusion matrix* [25]. Analisis sentimen yang digunakan pada penelitian dengan algoritma *naïve bayes* menggunakan data masa lampau yang dijadikan sebagai data pelatihan dan data masa depan sebagai data pengujian dengan rumus sebagai berikut [20]:

$$P(Ck|dj) = P(Ck) \prod_{i=1}^T P(ti|Ck) \quad (5)$$

Dimana $P(ti|Ck)$ adalah probabilitas bersyarat dari kata ti muncul dalam dokumen dari kelas C , sedangkan $P(Ck)$ adalah probabilitas sebelumnya dokumen terdapat dalam kelas Ck . Kedua probabilitas ini diestimasi dari data pelatihan [20].

L. Evaluasi

Data yang telah dilakukan pengujian, akan dievaluasi untuk melihat nilai keakuratan data tersebut. Indikator yang dilihat adalah data pada kolom *precision*, *recall* dan *f-1 score* dan yang paling penting adalah nilai *accuracy* [18], [26].

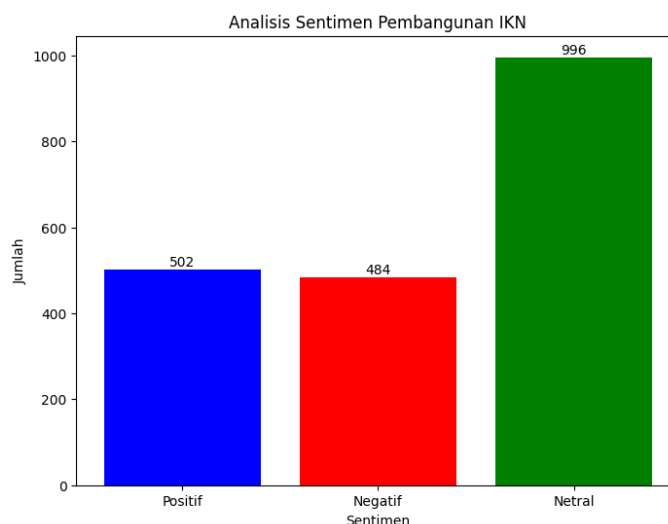
III. HASIL DAN PEMBAHASAN

Data hasil *crawling* dilakukan *pre-processing* data untuk mendapatkan kalimat atau kata yang bebas dari kata pengganggu dimulai dari proses *case folding*, *tokenizing*, *normalization*, *stemming* dan *filtering* sesuai Gambar 11.

	full_text	Text Case Folding	Text Tokenizing	Text Normalization	Text Stemming	Text Filtering
0	@__LOVE_AG4EVER @jokowi Jadi gak nih 17an di ikn	loveageever jadi gak nih an di ikn	[loveageever, jadi, gak, nih, an, di, ikn]	[loveageever, jadi, enggak, nih, an, di, ikn]	[loveageever, jadi, enggak, nih, an, di, ikn]	[loveageever, an, ikn]
1	@jeffsatur_Italy Like I always proud that he w...	italy like always proud that he will ready to...	[italy, like, always, proud, that, he, will, r...]	[italy, like, always, proud, that, he, will, r...]	[italy, like, always, proud, that, he, will, r...]	[italy, like, always, proud, that, he, will, r...]
2	Perkembangan tol IKN Ganta Bali Cantik Gaji 50...	perkembangan tol ikn ganta bali cantik gaji jl...	[perkembangan, tol, ikn, ganta, bali, cantik, ...]	[perkembangan, tol, ikn, ganta, bali, cantik, ...]	[kembang, tol, ikn, ganta, bal, cantik, gaji, ...]	[kembang, tol, ikn, ganta, bal, cantik, gaji, ...]
3	Bandara VVIP IKN Beroperasi Fungsional 1 Agust...	bandara vvip ikn beroperasi fungsional agustus...	[bandara, vvip, ikn, beroperasi, fungsional, a...]	[bandara, vvip, ikn, beroperasi, fungsional, a...]	[bandara, vvip, ikn, operasi, fungsional, agus...]	[bandara, vvip, ikn, operasi, fungsional, agus...]
4	mas @gibran_tweet ini area lahan pertanian di ...	mas tweet ini area lahan pertanian di ikn posi...	[mas, tweet, ini, area, lahan, pertanian, di, ...]	[mas, tweet, ini, area, lahan, pertanian, di, ...]	[mas, tweet, ini, area, lahan, tani, di, ikn, ...]	[mas, tweet, area, lahan, tani, ikn, posisi, d...]

Gambar 11. Hasil *Pre-processing* Data

Hasil dari *crawling* data dari media sosial X kemudian dilakukan *pre-processing data* yang hasilnya berupa data *text* yang sudah *clear*, kemudian dilakukan pelabelan dengan tujuan mengelompokkan data berdasarkan kelompok negatif, positif dan netral. Pelabelan atas opini atau sentimen ini dilakukan dengan cara otomatis menggunakan Inset Lexicon Indonesia yang menghasilkan nilai *polarity score*, bila nilai *polarity score* > 0 maka masuk ke dalam klasifikasi positif, bila *polarity score* < 0 maka akan masuk dalam kategori negatif dan apabila nilai *polarity score* = 0 maka masuk kedalam kategori netral. Dengan persebaran data seperti pada Gambar 12 yaitu data positif sejumlah 502 data, data negatif sejumlah 484 data dan data netral sejumlah 996 data dari keseluruhan 1982 data.



Gambar 12. Persebaran Data dalam Tiga Kategori

Data tersebut kemudian dipisah menjadi data pelatihan dan data pengujian dengan persentase data pelatihan sebesar 80% dan data pengujian sebesar 20%. Selanjutnya dilakukan pembobotan menggunakan vektor *Term Frequency-Inverse Document Frequency* yaitu suatu proses perubahan kata menjadi numerik [21]. Kemudian data akan dilatih menggunakan tiga algoritma klasifikasi yaitu *support vector machine*, *logistic regression* dan *naïve bayes*. Pada awal penelitian ini untuk klasifikasi *SVM* digunakan pengklasifikasian tiga jenis *support vector machine*, yaitu *support vector machine linear* [27], *support vector machine polynomial* dan *support vector machine radial basis function (RBF)* untuk melihat jenis algoritma mana yang memiliki nilai akurasi yang paling baik. Hasil dari klasifikasi menggunakan tiga algoritma *SVM* dapat tertera pada Gambar 13 dengan nilai akurasi *SVM Linear* 79%, Gambar 14 dengan nilai akurasi *SVM Polynomial* 72% dan Gambar 15 nilai akurasi *SVM radial basis function* sebesar 80%.

	precision	recall	f1-score	support
Negative	0.73	0.63	0.68	93
Neutral	0.80	0.94	0.86	199
Positive	0.82	0.64	0.72	105
accuracy			0.79	397
macro avg	0.78	0.74	0.75	397
weighted avg	0.79	0.79	0.78	397

Gambar 13. Hasil Akurasi Model *SVM Linear*

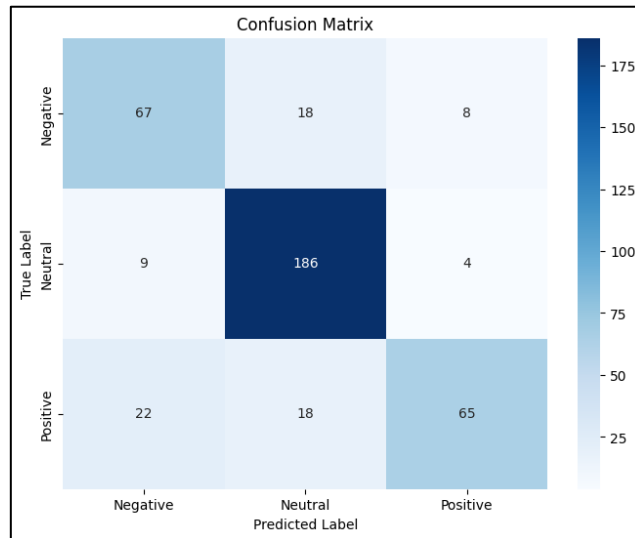
	precision	recall	f1-score	support
Negative	0.90	0.38	0.53	93
Neutral	0.64	0.98	0.78	199
Positive	0.94	0.49	0.64	105
accuracy			0.71	397
macro avg	0.83	0.62	0.65	397
weighted avg	0.78	0.71	0.68	397

Gambar 14. Hasil Akurasi Model *SVM Polynomial*

	precision	recall	f1-score	support
Negative	0.68	0.72	0.70	93
Neutral	0.84	0.93	0.88	199
Positive	0.84	0.62	0.71	105
accuracy			0.80	397
macro avg	0.79	0.76	0.77	397
weighted avg	0.80	0.80	0.80	397

Gambar 15. Hasil Akurasi Algoritma *SVM Radial Basis Function*

Dari hasil pengkalifikasian ketiga algoritma *SVM* tersebut dapat terlihat nilai akurasi *SVM linear* adalah 79%, nilai akurasi *SVM polynomial* adalah 71% dan nilai akurasi *SVM radial basis function* adalah 80%, maka dipilih *SVM radial basis function (RBF)* dengan memiliki nilai akurasi yang paling baik yaitu 80%.

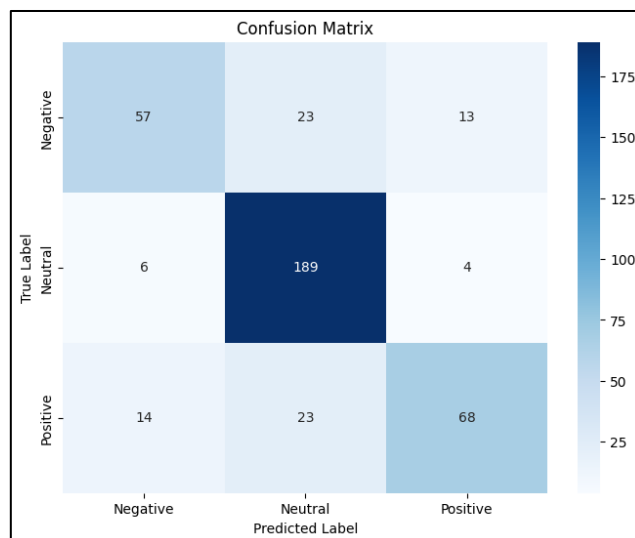


Gambar 16. Confusion Matrix SVM Radial Basis Function

Pada Gambar 16 dapat diketahui bahwa didapat prediksi benar pada klasifikasi positif sejumlah 65 data, klasifikasi negatif sebanyak 67 data dan klasifikasi netral sejumlah 186 data. Selanjutnya dilakukan pemodelan kedua dengan menggunakan algoritma *logistic regression*. Hasil pengklasifikasian menggunakan algoritma *logistic regression* seperti terlihat pada Gambar 17 dengan nilai akurasi 79%.

	precision	recall	f1-score	support
Negative	0.74	0.61	0.67	93
Neutral	0.80	0.95	0.87	199
Positive	0.80	0.65	0.72	105
accuracy			0.79	397
macro avg	0.78	0.74	0.75	397
weighted avg	0.79	0.79	0.78	397

Gambar 17. Hasil Akurasi Model Logistic Regression

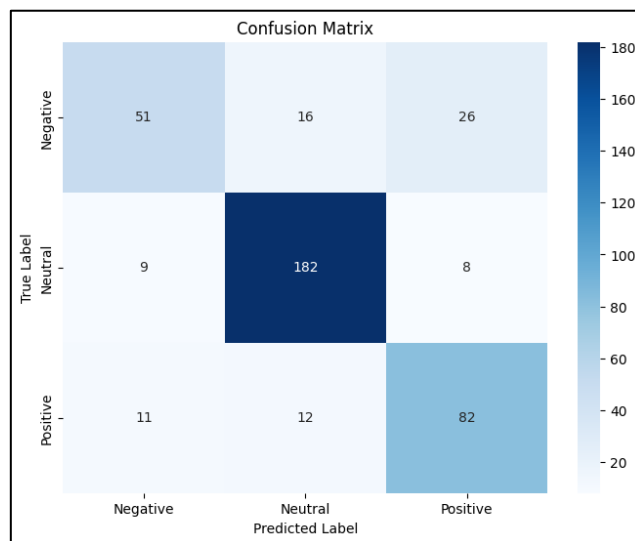


Gambar 18. Confuse Matrix Model Logistic regression

Pada Gambar 18 dapat diketahui bahwa didapat prediksi benar pada klasifikasi positif sebanyak 68 data, klasifikasi negatif sebanyak 57 data dan klasifikasi netral sebanyak 189 data. Selanjutnya dilakukan pemodelan ketiga dengan menggunakan algoritma *naïve bayes*. Hasil pengklasifikasian dengan algoritma *naïve bayes* terlihat pada Gambar 19 dengan nilai akurasi 79%.

	precision	recall	f1-score	support
Negative	0.72	0.55	0.62	93
Neutral	0.87	0.91	0.89	199
Positive	0.71	0.78	0.74	105
accuracy			0.79	397
macro avg	0.76	0.75	0.75	397
weighted avg	0.79	0.79	0.79	397

Gambar 19. Hasil Akurasi Algoritma *Naïve Bayes*



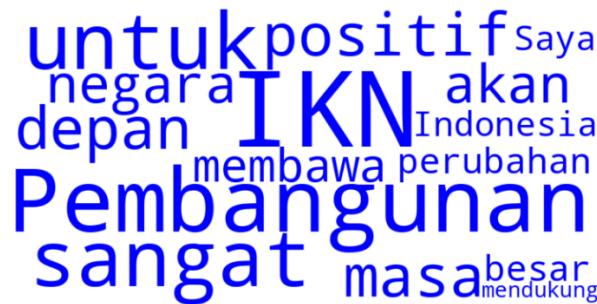
Gambar 20. Confuse Matrix *Naïve Bayes*

Pada Gambar 20 dapat diketahui bahwa didapat prediksi benar pada klasifikasi positif sebanyak 82 data, klasifikasi negatif sebanyak 51 data dan klasifikasi netral sebanyak 182 data. Dari pemodelan tiga algoritma ini maka dapat dilihat hasil akurasi data ujinya untuk *SVM radial basis function* yaitu 80 % dan akurasi data pengujian algoritma *logistic regression* dan *naïve bayes* yaitu 79%. Sedangkan untuk pengujian data latih sendiri bervariasi, untuk algoritma *support vector machine radial basis function* 99%, algoritma *logistic regression* 88%, algoritma *naïve bayes* 86% seperti terlihat pada Tabel 1.

TABEL 1
PERBANDINGAN KINERJA ALGORITMA

Algoritma	Data Latih (%)	Data Uji (%)
SVM RBF	99	80
Logistic Regression	88	79
Naïve Bayes	86	79

Setelah dilakukan perbandingan kinerja algoritma, selanjutnya, visualisasi data dilakukan menggunakan *word cloud* untuk setiap kategori sentimen, yaitu positif dan negatif. Gambar 21 menampilkan *word cloud* untuk kategori sentimen positif, sedangkan Gambar 22 menampilkan *word cloud* untuk kategori sentimen negatif. *Word cloud* ini memberikan representasi visual dari kata-kata yang paling sering muncul atau dominan di setiap kategori, sehingga dapat membantu dalam memahami pola sentimen secara lebih mendalam.



Gambar 21. Word Cloud positif



Gambar 22. Word Cloud negatif

IV. SIMPULAN

Hasil penelitian ini menunjukkan bahwa algoritma *Support Vector Machine (SVM)* dengan fungsi *Radial Basis Function (RBF)* memiliki akurasi terbaik dalam menganalisis sentimen masyarakat terkait pembangunan Ibu Kota Negara (IKN) di Pulau Kalimantan, dengan akurasi pengujian sebesar 80% dan akurasi pelatihan 99%, dibandingkan dengan algoritma *Logistic Regression* dan *Naïve Bayes* yang memiliki akurasi pengujian sebesar 79% dan akurasi pelatihan masing-masing 88% dan 86%. Hasil analisis sentimen menunjukkan bahwa pendapat masyarakat terkait pembangunan IKN cenderung seimbang antara sentimen positif dan negatif, dengan sentimen netral paling dominan akibat informasi dari akun resmi terkait progres pembangunan. Untuk penelitian selanjutnya, disarankan agar digunakan dataset yang lebih besar dengan periode pengambilan data yang lebih panjang atau kata kunci tambahan, guna memperoleh gambaran sentimen masyarakat yang lebih lengkap. Penelitian juga bisa melibatkan analisis temporal untuk mengidentifikasi tren sentimen seiring waktu, serta memanfaatkan model *deep learning* seperti *recurrent neural networks (RNN)* atau *transformers*, dengan penerapan metode *tuning hyperparameter* seperti *grid search* atau *random search* untuk meningkatkan akurasi dan kinerja model.

UCAPAN TERIMA KASIH

Terima kasih kami ucapkan kepada Program Studi Magister Teknik Industri Universitas Trisakti yang telah membimbing dan mengizinkan untuk mengajukan penulisan penelitian ini.

DAFTAR PUSTAKA

- [1] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [2] A. Dyah Masitah, D. Suluh, and K. Dewi, "Analisis Opini Publik Berdasarkan Teori Agenda Setting Pada Proses Perencanaan Pemindahan IKN," *Jurnal Ilmu Sosial dan Pendidikan (JISIP)*, vol. 6, no. 3, pp. 2598–9944, 2022, doi: 10.36312/jisip.v6i3.3374/http.
- [3] G. Aji, Z. Arfani, A. M. Sari, R. Septiani, and U. K. H. Abdurrahman Wahid, "Dampak Pemindahan Ibukota Negara Baru terhadap Ekonomi dan Sosial di Provinsi Kalimantan Timur," 2023. [Online]. Available: <http://jurnal.kolibi.org/index.php/kultura>
- [4] D. Ocsanda, C. Ilham Wijaya, M. Azzam Al Haq, J. Dwi Efendi, and F. Prihantoro, "OPORTUNITAS PEMBENTUKAN ENTITAS KEBUDAYAAN BARU, TINJAUAN TERHADAP IBU KOTA NEGARA (IKN) INDONESIA 2024 BERDASARKAN SEJARAH PERPINDAHAN IBU KOTA VOC 1619 OPPORTUNITY FOR NEW CULTURAL ENTITY EMERGENCE, AN OVERVIEW OF THE CAPITAL CITY (IKN) OF INDONESIA 2024 BASED ON THE HISTORY OF THE RELOCATION OF THE CAPITAL CITY OF VOC 1619," vol. 18, pp. 1–12, 2023, doi: 10.47441/jkp.v18i1.291.
- [5] M. V. Santos, F. Morgado-Dias, and T. C. Silva, "Oil Sector and Sentiment Analysis—A Review," Jun. 01, 2023, *MDPI*. doi: 10.3390/en16124824.
- [6] M. Amiruddin Saddam, E. D. Kurniawan, F. Teknologi Informasi, U. Budi Luhur, and J. Ciledug Raya, "Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine," vol. 8, no. 3, 2023.
- [7] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," *Journal of Information and Telecommunication*, vol. 5, no. 1, pp. 1–15, 2021, doi: 10.1080/24751839.2020.1790793.

-
- [8] Z. Kastrati, F. Dalipi, A. S. Imran, K. P. Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study," 2021, *MDPI AG*. doi: 10.3390/app11093986.
- [9] K. Garcia and L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA," *Appl Soft Comput*, vol. 101, Mar. 2021, doi: 10.1016/j.asoc.2020.107057.
- [10] C. Humam and A. D. Laksito, "Implementasi Aplikasi Sentimen Pada Data Twitter Jelang Pemilu 2024," vol. 8, no. 2, 2023.
- [11] A. N. Syafia, M. F. Hidayattullah, and W. Sutteddy, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," vol. 8, no. 3, 2023.
- [12] C. Humam and A. D. Laksito, "Implementasi Aplikasi Sentimen Pada Data Twitter Jelang Pemilu 2024," vol. 8, no. 2, 2023.
- [13] C. Ayunda *et al.*, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," vol. 7, no. 2, 2022.
- [14] R. B. Dahlian and D. Sitanggang, "Sentiment Analysis of Digital Television Migration on Twitter Using Naïve Bayes Multinomial Comparison, Support Vector Machines, and Logistic Regression Algorithms," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 280–288, Jul. 2023, doi: 10.32736/sisfokom.v12i2.1668.
- [15] A. Rahman Isnain, N. Hendrastuty, and L. Andraini, "Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis," vol. 6, no. 1, 2021.
- [16] "SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE BASED ON FEATURE SELECTION AND SEMANTIC ANALYSIS," *International Research Journal of Computer Science (IRJCS) Citation: Dr .Ar ivoli & Sonali*, pp. 209–214, 2021, doi: 10.26562/ir.
- [17] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.
- [18] I. Habib Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, 2023.
- [19] ... | Sumertajaya, I. M. Angraini, Y. R. Harahap, and J. B. Fitrianto, "Sentiment Analysis on Covid-19 Vaccination in Indonesia Using Support Vector Machine and Random Forest." [Online]. Available: <https://apps.twitter.com/>
- [20] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 3, no. 3, pp. 394–401, Oct. 2018, doi: 10.30591/jpit.v3i3.977.
- [21] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.
- [22] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.
- [23] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/5594899.
- [24] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, "COMPARISON OF LOGISTIC REGRESSION, MULTINOMIALNB, SVM, AND K-NN METHODS ON SENTIMENT ANALYSIS OF GOJEK APP REVIEWS ON THE GOOGLE PLAY STORE," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1487–1494, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.863.
- [25] K. Nurbagja *et al.*, "Sentiment Analysis of the Increase in Fuel Prices Using Random Forest Classifier Method," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 1, pp. 132–144, 2023, doi: 10.12928/biste.v5i1.7414.
- [26] H. Kaur, S. Ul Ahsaan, B. Alankar, and V. Chang, "A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets", doi: 10.1007/s10796-021-10135-7/Published.
- [27] A. A. Aldino, A. Saputra, A. Nurkholis, and S. Setiawansyah, "Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 325–330, Dec. 2021, doi: 10.47065/bits.v3i3.1041.