

Analisa Sentimen Rencana Pemindahan Ibu Kota Nusantara dari Jakarta ke Kalimantan Timur Menggunakan Algoritma Naïve Bayes

Im Muhaemin Abdul Azis ¹, Tri Wahyudi ^{2*}

^{1,2*} Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: iimmuhaemin38@gmail.com ¹, triwahyudi100390@gmail.com ^{2*}

Histori Artikel:

Dikirim 25 Mei 2024; Diterima dalam bentuk revisi 20 Juni 2024; Diterima 20 Juli 2024; Diterbitkan 20 September 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Pemindahan pusat pemerintahan atau ibu kota negara Indonesia telah menjadi topik yang sangat diperbincangkan dalam beberapa tahun terakhir. Keputusan ini menimbulkan beragam pandangan dan pendapat masyarakat, terutama yang tercermin melalui media sosial Twitter. Penelitian ini bertujuan memahami pandangan serta sikap masyarakat Indonesia terkait rencana pemindahan ibu kota negara ke Nusantara yang tercermin di Twitter. Metode analisis sentimen menggunakan algoritma Naïve Bayes digunakan untuk mengklasifikasikan teks-teks dari Twitter ke dalam kategori sentimen positif dan negatif. Dengan data dari Twitter, diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam tentang persepsi masyarakat terhadap pemindahan pusat pemerintahan. Dari hasil penelitian, total dataset yang digunakan berjumlah 375 data. Hasil perhitungan menggunakan RapidMiner menunjukkan precision untuk sentimen positif adalah 94.89% dan untuk sentimen negatif 65.00%. Recall untuk sentimen negatif adalah 48.15%, sementara recall untuk sentimen positif mencapai 97.38%. Nilai akurasi yang dihasilkan oleh model adalah 92.82%, menunjukkan bahwa algoritma Naïve Bayes dapat mengklasifikasikan sentimen dengan baik menggunakan data yang ada. Penelitian ini memberikan gambaran tentang dukungan dan penolakan masyarakat serta faktor-faktor yang memengaruhi sikap mereka terhadap pemindahan ibu kota.

Kata Kunci: Ibu Kota; Twitter; Data Mining; Text Mining; Analisa Sentimen; Naïve Bayes.

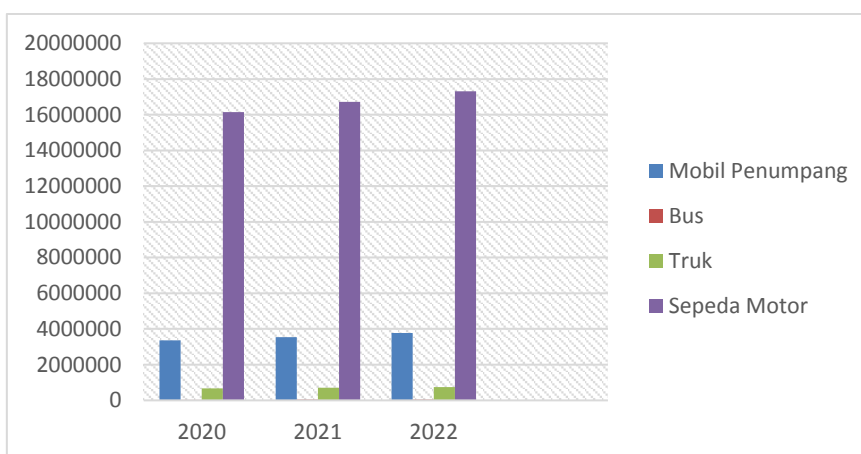
Abstract

The relocation of Indonesia's capital city has been a highly debated topic in recent years. This decision has sparked various opinions and views from the public, especially as reflected on social media platforms like Twitter. This research aims to understand the perspectives and attitudes of Indonesian citizens regarding the plan to move the capital to Nusantara, as expressed on Twitter. Sentiment analysis using the Naïve Bayes algorithm is employed to classify tweets into positive and negative sentiment categories. By leveraging data from Twitter, this study aims to provide deeper insights into public perception and attitudes towards the relocation of the government center. The research used a dataset of 375 tweets. The results, analyzed using RapidMiner, show that the precision for positive sentiment is 94.89% and for negative sentiment is 65.00%. The recall for negative sentiment is 48.15%, while the recall for positive sentiment reaches 97.38%. The model's accuracy is 92.82%, indicating that the Naïve Bayes algorithm can effectively classify sentiments using the available data. This study offers insights into public support and opposition, as well as the factors influencing their attitudes towards the capital relocation plan.

Keyword: Capital City; Twitter; Data Mining; Text Mining; Sentiment Analyst; Naïve Bayes.

1. Pendahuluan

Sejarah pemindahan ibu kota negara Indonesia telah memiliki rekam jejak yang panjang. Sebelumnya, perpindahan ibu kota bukanlah hal asing dalam sejarah Indonesia. Sebagai contoh, sebelum Jakarta menjadi ibu kota negara pada tahun 1949, ibu kota Indonesia pertama kali berada di Kota Yogyakarta pada masa pendudukan Belanda. Setelah proklamasi kemerdekaan pada tahun 1945, Yogyakarta tetap menjadi pusat pemerintahan Republik Indonesia hingga pemindahan ibu kota ke Jakarta pada tahun 1950. Keputusan ini diambil karena Jakarta memiliki keuntungan strategis sebagai ibu kota yang terletak di tengah-tengah kepulauan Indonesia. Namun, seiring berjalannya waktu, Jakarta mengalami berbagai masalah seperti kemacetan lalu lintas, banjir, dan penurunan permukaan tanah. Data kemacetan di Jakarta pada tahun 2022 menurut sumber Korlantas POLRI menunjukkan bahwa jumlah kendaraan bermotor terus meningkat dari tahun ke tahun, dengan total jumlah kendaraan bermotor di Provinsi DKI Jakarta terus meningkat dari tahun 2020 hingga 2022. Pada tahun 2020, terdapat total 20.221.821 unit kendaraan bermotor, yang terdiri dari 3.365.467 mobil penumpang, 35.266 bus, 679.708 truk, dan 16.141.380 sepeda motor. Pada tahun 2021, jumlah kendaraan meningkat menjadi 21.005.527 unit, dengan peningkatan terutama terjadi pada mobil penumpang, bus, dan truk. Pada tahun 2022, jumlah kendaraan bermotor mencapai 21.856.081 unit, dengan peningkatan terutama terjadi pada semua jenis kendaraan, termasuk mobil penumpang yang mencapai 3.766.059 unit, bus sebanyak 37.180 unit, truk sebanyak 748.395 unit, dan sepeda motor sebanyak 17.304.447 unit.



Gambar 1. Jumlah Kendaraan Bermotor Menurut Jenis Kendaraan (unit) di Provinsi DKI Jakarta (Sumber : Korlantas POLRI)

Badan Pusat Statistik (BPS) juga menyatakan bahwa Pulau Jawa merupakan penyumbang terbesar terhadap perekonomian Indonesia, memberikan kontribusi sebesar 57,05% terhadap produk domestik bruto (PDB). Sumatera berada di posisi kedua dengan andil sebesar 22,01%, diikuti oleh Kalimantan dengan 8,49%, Sulawesi dengan 7,10%, Bali dan Nusa Tenggara dengan 2,77%, serta Maluku dan Papua dengan kontribusi sebesar 2,58%. Selain itu, BPS mengungkapkan bahwa pada tahun 2023, seluruh kelompok pulau menunjukkan pertumbuhan ekonomi yang baik meskipun ada tantangan dari pelemahan ekonomi global. Pertumbuhan ekonomi tertinggi terjadi di Maluku dan Papua sebesar 6,94%, diikuti oleh Sulawesi dengan 6,37%. Kalimantan mengalami pertumbuhan sebesar 5,43%, Jawa 4,96%, Sumatera 4,69%, dan pertumbuhan terendah tercatat di Bali dan Nusa Tenggara dengan 4%. Hal ini mendorong kebutuhan akan pemindahan ibu kota untuk mengurangi beban di Jakarta dan mengalihkan pembangunan lebih merata ke wilayah Indonesia timur. Keputusan pemerintah Indonesia untuk memindahkan ibu kota ke Nusantara adalah sebuah langkah strategis yang bertujuan menciptakan keseimbangan pembangunan di seluruh wilayah negara. Namun, keputusan ini juga menjadi subjek perdebatan kompleks di masyarakat. Dalam konteks ini, analisis

sentimen menjadi sangat penting untuk memahami sikap masyarakat terhadap rencana tersebut, termasuk dukungan dan penolakan, serta faktor-faktor yang memengaruhi sikap mereka. Analisis sentimen merupakan bagian dari penelitian dalam text mining, yang terhubung dengan bidang lebih luas seperti pemrosesan data untuk kegiatan spesifik (Darwis *et al.*, 2021). Terdapat banyak algoritma yang dapat digunakan untuk melakukan analisis sentimen, salah satu algoritma yang sering digunakan adalah Naïve Bayes. Meskipun sederhana, algoritma ini efisien dalam mengklasifikasikan sentimen berdasarkan teorema Bayes dengan asumsi fitur independen, sehingga memberikan pemahaman yang berharga dalam konteks perdebatan publik, seperti pemindahan ibu kota (Hasri & Alita, 2022).

Sudah banyak penelitian sebelumnya yang menggunakan algoritma Naïve Bayes sebagai metode untuk melakukan analisis sentimen. Seperti pada peneliti yang dilakukan oleh Safitri Juanita dengan judul “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes”. Dari hasil penelitian tersebut didapatkan hasil menunjukkan bahwa dari dataset tweet pemilu 2019 dari periode 16 April 2018 hingga 16 April 2019 cenderung memiliki pola persepsi negatif yang dominan, mencapai 52%, sementara persepsi positif hanya sebesar 18%, dan persepsi netral sebesar 31%. Meskipun demikian, tingkat akurasi klasifikasi Naive Bayes terhadap dataset training mencapai 81%, sementara untuk dataset testing mencapai 76%. Nilai rata-rata precision untuk sentimen positif, negatif, dan netral masing-masing adalah 86,65%, 77,15%, dan 80,95%, dengan nilai recall untuk sentimen positif sebesar 36,8%, sentimen negatif sebesar 93,2%, dan sentimen netral sebesar 86,8% (Juanita, 2020). Pada penelitian lain yang dilakukan oleh Vynska Amalia Permadi dengan judul “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura”. Dari penelitian tersebut menunjukkan bahwa pengelompokan review kepuasan pengunjung menggunakan algoritma Naïve Bayes menghasilkan akurasi sebesar 73%. Visualisasi hasil klasifikasi ditampilkan dalam aplikasi berbasis web menggunakan R Shiny, yang menyajikan hasil analisis dalam bentuk wordcloud dan diagram. Dengan demikian, penelitian ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan secara efektif untuk mengklasifikasikan sentimen komentar pengunjung restoran, membantu pengunjung lain dalam membuat keputusan yang lebih baik. (Permadi, 2020)

Berdasarkan latar belakang diatas, penulis tertarik untuk membuat penelitian dengan judul “Analisa Sentimen Rencana Pemindahan Ibu Kota Nusantara Dari Jakarta Ke Kalimantan Timur Menggunakan Algoritma Naive Bayes”. Penelitian ini bermaksud untuk mengetahui pandangan masyarakat Indonesia terhadap rencana pemindahan ibu kota negara pada media sosial twitter. Penelitian ini menggunakan algoritma Naïve Bayes yang digunakan untuk mengklasifikasikan teks atau dokumen ke dalam kategori sentimen tertentu, seperti positif dan negatif berdasarkan fitur-fitur atau kata-kata yang terdapat dalam teks tersebut. Naive Bayes dianggap sebagai metode yang lebih unggul untuk pemisahan data terstruktur dibandingkan metode lainnya, baik dalam hal akurasi maupun efisiensi komputasi (Ananda & Suryono, 2024).

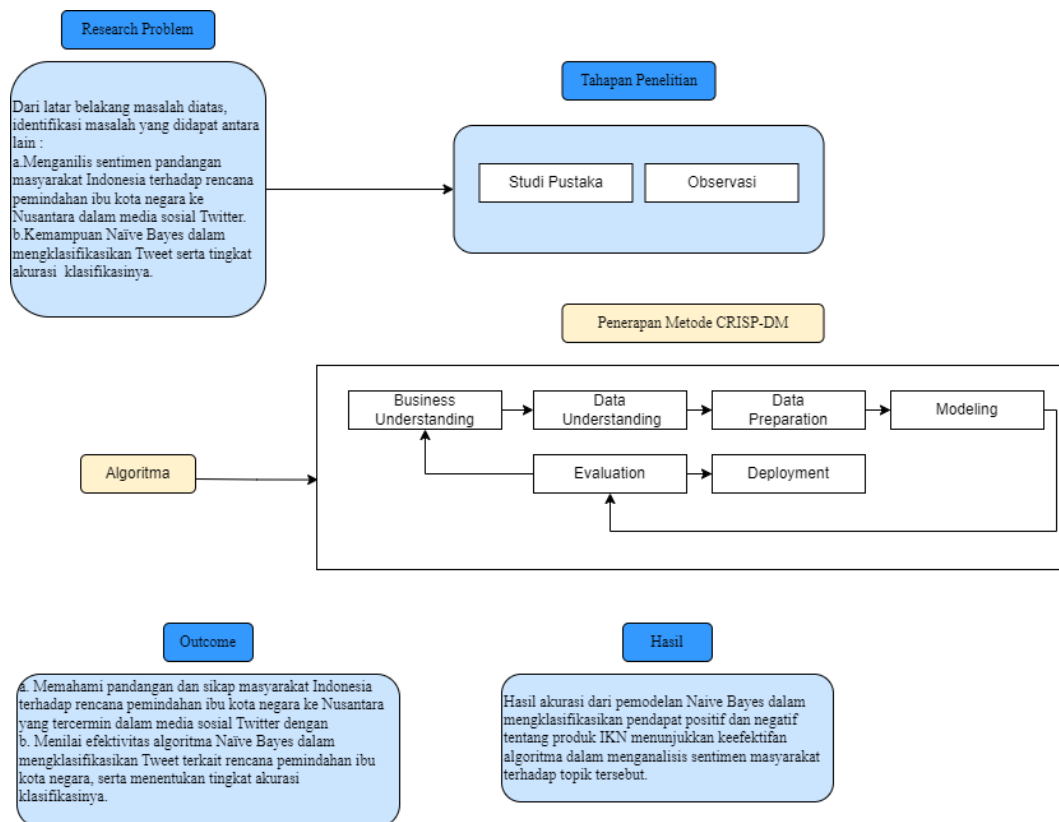
2. Metode Penelitian

2.1 Penerapan Metodologi

Metodologi yang digunakan dalam penelitian ini melibatkan pengujian terhadap sentimen publik di Twitter mengenai rencana pemindahan ibu kota Nusantara dengan menggunakan algoritma Naïve Bayes yang diimplementasikan melalui perangkat lunak RapidMiner. RapidMiner dipilih karena kemampuannya dalam menangani data berukuran besar secara efisien. Alat ini memanfaatkan berbagai operator untuk mengelola dan memproses data secara otomatis. Dalam konteks penelitian ini, data Twitter dihubungkan dengan node operator, yang selanjutnya mengarahkan data tersebut ke node hasil untuk menampilkan output yang diinginkan (Ainurrohma, 2021). Langkah pertama dalam proses ini adalah pengumpulan data, yang dilakukan secara otomatis melalui skrip yang dijalankan di platform Google Colab. Skrip ini ditulis dalam bahasa pemrograman Python, yang dikenal karena kemudahannya dalam digunakan di berbagai platform dan kemampuannya dalam menghasilkan kode

yang mudah dipahami (Alfandi Safira & Hasan, 2023). Setelah data dikumpulkan, tahap selanjutnya adalah pembersihan data. Proses ini melibatkan penghapusan elemen-elemen yang tidak relevan, seperti mention, hashtag, serta simbol-simbol lain yang tidak berkontribusi terhadap analisis sentimen. Setelah itu, data melewati beberapa tahap preprocessing yang meliputi tokenisasi (pemecahan teks menjadi unit-unit kecil seperti kata), normalisasi (penyederhanaan teks), dan penghapusan *stop words*, yaitu kata-kata umum yang tidak memiliki makna penting dalam konteks analisis, seperti "dan," "atau," dan "tetapi".

Tahap berikutnya dalam metodologi ini adalah penerapan *Cross-Industry Standard Process for Data Mining* (CRISP-DM), sebuah model yang diakui secara luas dalam proses data mining. CRISP-DM menyediakan kerangka kerja standar yang dapat diterapkan dalam berbagai konteks analisis data, dan membantu peneliti dalam menguraikan permasalahan dengan pendekatan yang sistematis. Metodologi ini mencakup enam tahapan utama, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment (Rahman Hakim & Sugiyono, 2024). Dalam penelitian ini, tahapan-tahapan ini digunakan untuk mengolah data Twitter terkait rencana pemindahan ibu kota, sehingga menghasilkan analisis yang akurat dan dapat diandalkan mengenai sentimen publik.



Gambar 2. Gambar Tahapan Penerapan Metodologi

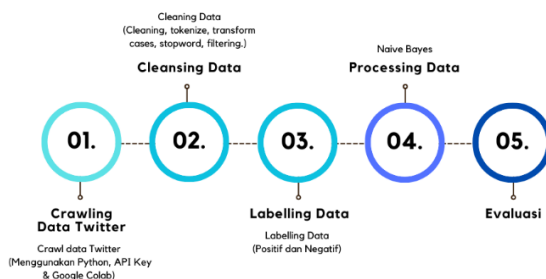
Pada gambar di atas, terdapat enam tahapan dalam proses CRISP-DM yang dapat dijelaskan sebagai berikut. Pertama, Pemahaman Bisnis (*Business Understanding*), yaitu tahap yang bertujuan untuk memahami tujuan penelitian dari perspektif ilmiah. Fokus utama tahap ini adalah merumuskan masalah penelitian secara jelas dan spesifik serta menetapkan kriteria keberhasilan yang sesuai dengan tujuan penelitian. Kedua, Pemahaman Data (*Data Understanding*), di mana data awal dikumpulkan dan dianalisis untuk memahami isinya, mengidentifikasi masalah kualitas data, serta mendapatkan wawasan awal. Eksplorasi data dilakukan untuk menemukan pola atau anomali yang relevan dengan hipotesis penelitian.

Tahap ketiga adalah Data Preparation yang mencakup semua aktivitas yang diperlukan untuk membentuk dataset akhir. Ini melibatkan seleksi, pembersihan, transformasi, dan integrasi data agar siap digunakan dalam analisis. Pengumpulan data dimulai dengan mengumpulkan tweet yang menggunakan kata kunci "Ibu Kota Nusantara." Pembersihan data dilakukan dengan mengubah semua teks menjadi huruf kecil, menghapus tanda baca yang tidak relevan, serta menghilangkan stop words seperti "dan," "atau," dan "tapi," untuk fokus pada kata yang lebih bermakna. Selanjutnya, dilakukan tokenisasi, yaitu proses memecah teks menjadi unit-unit kecil yang disebut token. Token ini kemudian digunakan sebagai input untuk model pembelajaran mesin (Naraswati *et al.*, 2021). Setelah itu, data diberikan label dalam dua kategori: positif dan negatif, berdasarkan karakteristik teks seperti komentar yang berisi kata-kata positif atau negatif.

Tahap keempat adalah Modelling, di mana dilakukan pemilihan dan penerapan teknik pemodelan yang sesuai dengan pertanyaan penelitian. Dalam konteks penelitian ini, teknik analisis sentimen menggunakan algoritma Naïve Bayes digunakan untuk mengklasifikasikan sentimen positif dan negatif dari tweet. Tahap kelima adalah Evaluation, di mana model yang telah dibangun dievaluasi secara menyeluruh untuk memastikan validitas dan reliabilitasnya. Evaluasi ini bertujuan untuk menentukan apakah model yang digunakan memenuhi kriteria penelitian. Terakhir, Deployment, yang merupakan tahap penyusunan laporan hasil kegiatan data mining. Laporan ini tidak hanya mencakup pengetahuan yang diperoleh, tetapi juga pengenalan pola-pola signifikan yang muncul selama proses analisis data.

2.2 Proses Pengumpulan Data

Data tersebut diperoleh dari twitter dengan menggunakan skrip otomatis untuk Twitter dengan kata kunci Ibu Kota Nusantara (IKN). Data yang berhasil diperoleh sebanyak 560 data. Berikut merupakan tahapan penelitian:



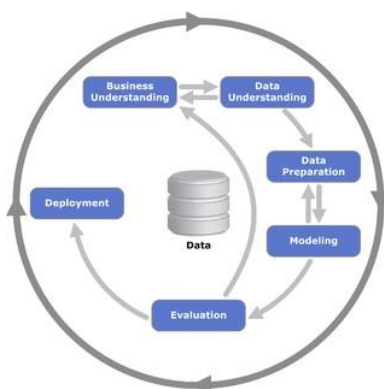
Gambar 3. Tahapan Penelitian

Penelitian ini melibatkan lima proses utama. Pertama, *Crawling* Data Twitter, di mana data diambil dari opini-opini yang terdapat di Twitter terkait rencana pemindahan ibu kota. Kata kunci yang digunakan adalah "Ibu Kota Nusantara," dengan tweet yang ditulis dalam bahasa Indonesia sebagai objek penelitian. Kedua, *Cleansing* Data, yaitu proses pembersihan data melalui beberapa tahap, termasuk *cleaning*, *tokenize*, transformasi huruf (*transform cases*), penghapusan *stop words*, dan penyaringan (*filtering*) untuk memastikan data siap digunakan. Selanjutnya, *Labelling* Data dilakukan pada data yang telah dibersihkan. Pelabelan awal dilakukan secara manual, dan kemudian model pelabelan otomatis dikembangkan untuk mengklasifikasikan data lebih lanjut. Setelah proses pelabelan, data diproses melalui tahap *Processing* Data, di mana pengujian dilakukan menggunakan metode pembobotan *TF-IDF*. Pengujian ini bertujuan untuk mengevaluasi model yang dibuat dengan Algoritma *Naïve Bayes*, dan untuk menentukan tingkat akurasi algoritma dalam mengklasifikasikan data. Tahap terakhir adalah *Evaluasi*, di mana hasil klasifikasi dianalisis, dan akurasi model diukur dengan menghitung persentase sentimen positif dan negatif dari data yang telah diproses.

3. Hasil dan Pembahasan

3.1 Hasil

Pada penelitian ini, data yang digunakan diperoleh dari media sosial Twitter menggunakan skrip otomatis untuk Twitter menggunakan Google Colab. Algoritma Naïve Bayes digunakan untuk mengklasifikasikan dataset yang diambil dari tweet Twitter. Proses pengolahan dan analisis data mengikuti metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) untuk memastikan pendekatan yang terstruktur dan sistematis. Aplikasi RapidMiner digunakan sebagai alat utama dalam pengolahan data, mulai dari pra-pemrosesan teks, penerapan model, hingga evaluasi hasil klasifikasi.



Gambar 4. Metodologi CRISP-DM

3.1.1 Business Understanding (Pemahaman Bisnis)

Keputusan pemerintah Indonesia untuk memindahkan ibu kota ke Nusantara adalah sebuah langkah strategis yang bertujuan menciptakan keseimbangan pembangunan di seluruh wilayah negara. Namun, keputusan ini juga menjadi subjek perdebatan kompleks di masyarakat. Dalam konteks ini, tujuan dasar adalah untuk menentukan mencari nilai akurasi, presisi, recall dan sentimen analisis pada dataset tentang pemindahan ibukota Indonesia.

3.1.2 Data Understanding (Pemahaman Data)

Pemahaman data dalam penelitian ini yaitu, mengumpulkan dan memahami data yang tersedia. Data didapat dari hasil skrip otomatis menggunakan Google Colab dengan menggunakan keyword Ibu Kota Nusantara. Proses ini dilakukan untuk mendapatkan data yang relevan dan mendukung dalam pemahaman data.

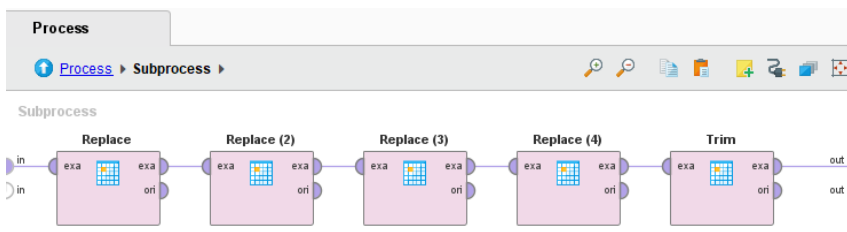
```

Crawl Tweet Twitter.ipynb
File Edit View Insert Runtime Tools Help Last edited on June 11
Code + Text
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keys/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keys/nodesource.gpg] https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo tee /etc/apt/sources
!sudo apt-get update
!sudo apt-get install nodejs -y
!node -v
# Crawl Data
filename = '1kn.csv'
search_keyword = 'ibu kota nusantara lang:id'
limit = 1500
!npx -y tweet-harvest@2.6.0 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
!import pandas as pd
# Specify the path to your CSV file
file_path = f'tweets-data/{filename}'
# Read the CSV file into a pandas DataFrame
    
```

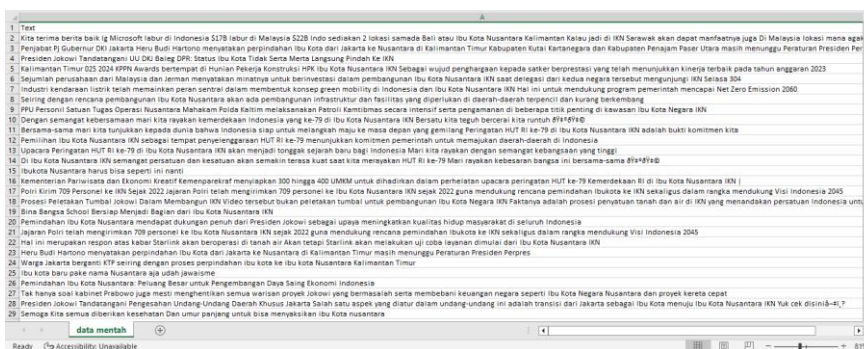
Gambar 5. Proses Crawling Data

3.1.3 Data Preparation (Persiapan Data)

Data Preparation pada penelitian ini yaitu memproses dan mempersiapkan data untuk analisis. Dalam mempersiapkan data diperlukan proses pembersihan data yaitu, mengubah semua huruf dalam teks menjadi huruf kecil untuk menghindari perbedaan antara huruf besar dan kecil. Menghapus tanda baca yang tidak berkontribusi pada analisis sentimen. Serta penghapusan *Stop Words*, *Stopwords* dapat diartikan sebagai menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan (seperti "dan", "atau", "tapi") untuk fokus pada kata-kata yang lebih bermakna (Kusuma & Nugroho, 2021).

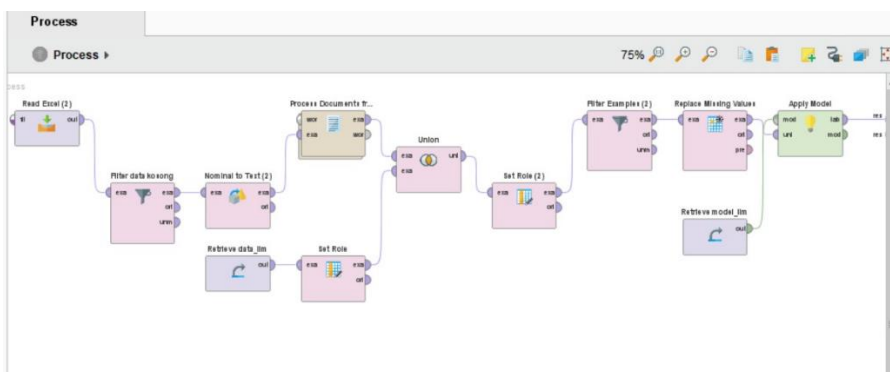


Gambar 6. Proses Pembersihan Data

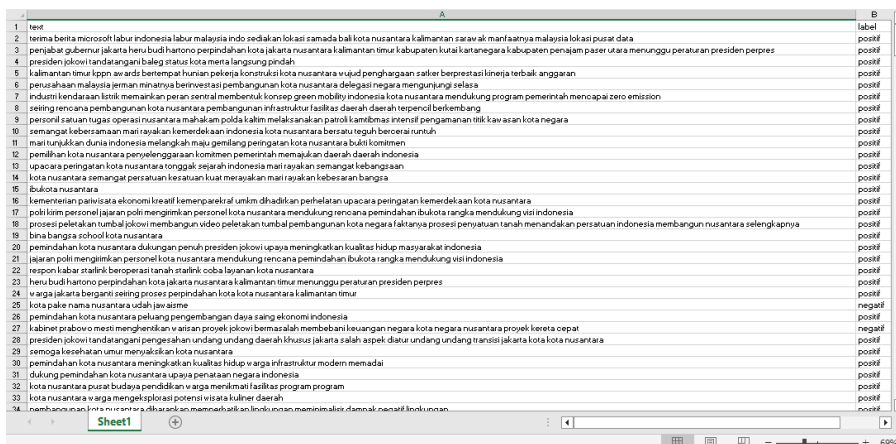


Gambar 7. Hasil Pembersihan Data

Dari hasil pembersihan kata, diperoleh data bersih sebanyak 375 data. Selanjutnya, dilakukan pelabelan dengan mengkategorikan data ke dalam dua kelas: positif dan negatif. Sebanyak 100 data digunakan sebagai data latih untuk memberi label sentimen secara manual. Proses pelabelan ini kemudian digunakan untuk melatih model, yang selanjutnya secara otomatis mengkategorikan sisa data.



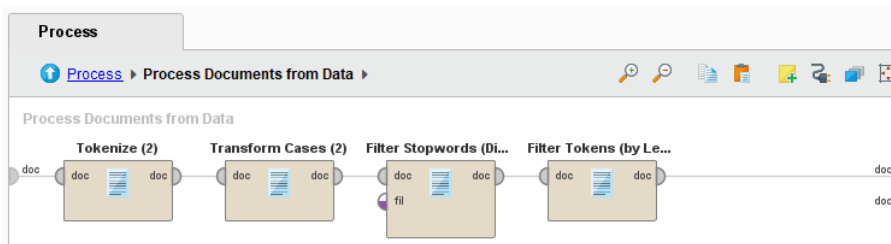
Gambar 8. Proses Labelling



Gambar 9. Hasil Labelling

3.1.4 Modelling

Modelling pada penelitian ini yaitu dilakukan *Preprocessing* Data yang bertujuan untuk mengubah data mentah menjadi data yang siap diolah. Pada proses ini juga untuk menghitung pembobotan kata untuk keperluan pada proses modelling.

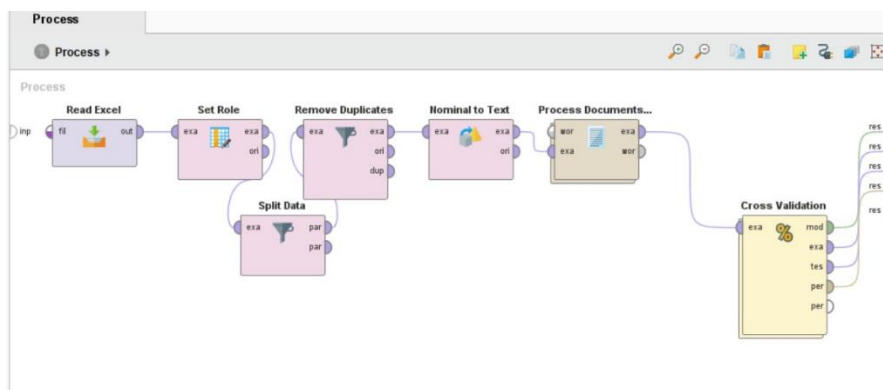


Gambar 10. Preprocessing Data

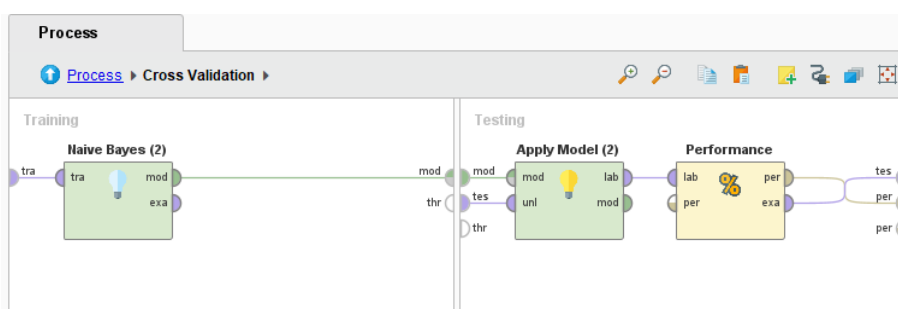
Row No.	label	text	abadikan	abdullah	acuannya	adhi	administrasi	advokasi	agama	agenda	agraria
1	posiftr	presiden joko...	0	0	0	0	0	0	0	0	0
2	posiftr	kalimantan t...	0	0	0	0	0	0	0	0	0
3	posiftr	perusahaan ...	0	0	0	0	0	0	0	0	0
4	posiftr	industri kend...	0	0	0	0	0	0	0	0	0
5	posiftr	seiring renca...	0	0	0	0	0	0	0	0	0
6	posiftr	personil satu...	0	0	0	0	0	0	0	0	0
7	posiftr	semangat ke...	0	0	0	0	0	0	0	0	0
8	posiftr	pemilihan kot...	0	0	0	0	0	0	0	0	0
9	posiftr	kota nusantar...	0	0	0	0	0	0	0	0	0
10	posiftr	polri kirim per...	0	0	0	0	0	0	0	0	0
11	posiftr	bina bangsa ...	0	0	0	0	0	0	0	0	0
12	posiftr	pemindahan ...	0	0	0	0	0	0	0	0	0
13	posiftr	jajaran poli...	0	0	0	0	0	0	0	0	0
14	posiftr	respon kabar...	0	0	0	0	0	0	0	0	0
15	posiftr	heru budi har...	0	0	0	0	0	0	0	0	0
16	negatif	kota pake na...	0	0	0	0	0	0	0	0	0
17	posiftr	pemindahan ...	0	0	0	0	0	0	0	0	0

Gambar 11. Hasil Tokenizing Data

Selanjutnya, pada tahap *modelling* peneliti akan melakukan pengukuran performa klasifikasi dengan menggunakan split data. Split data dibagi menjadi 2 yaitu 0.7 dan 0.3. Berikut proses *modelling* yang dilakukan pada RapidMiner ditunjukkan pada gambar berikut.



Gambar 12. Proses Modelling



Gambar 13. Proses Pengujian

3.1.5 Evaluation

Pada tahap ini akan dilakukan pengujian data dari tahapan modelling yang telah dibuat sebelumnya dengan menggunakan algoritma Naive Bayes. Total dataset yang dipakai yaitu berjumlah 375 data. Berikut adalah hasil dari tahapan modelling yang dapat dilihat dari hasil perhitungan hasil RapidMiner yang ditunjukkan pada gambar berikut.

accuracy: 92.82% +/- 3.84% (micro average: 92.86%)

	true positif	true negatif	class precision
pred. positif	260	14	94.89%
pred. negatif	7	13	65.00%
class recall	97.38%	48.15%	

Gambar 14. Hasil Proses Akurasi

Berdasarkan pada gambar diatas, dapat dibuat kesimpulan bahwa precision menunjukkan tingkat ketepatan data yang diprediksi positif terhadap banyaknya data yang benar diprediksi positif, yang menghasilkan persentase ketepatannya adalah 94.89%. Sedangkan untuk data bersentimen negatif, precision memiliki nilai sebesar 65.00%. Recall (*Specificity*) untuk sentimen negatif adalah 48.15%, sehingga dapat disimpulkan bahwa model dapat menemukan kembali informasi atau data yang benar-benar negatif dengan cukup baik. Sementara itu, recall untuk sentimen positif mencapai 97.38%. Nilai accuracy yang dihasilkan oleh model adalah 92.82%, sehingga dapat disimpulkan bahwa algoritma yang digunakan dapat mengklasifikasikan sentimen dengan baik menggunakan data yang ada.



Gambar 15. Wordcloud IKN

Wordcloud di atas menggambarkan kata-kata yang paling sering muncul dalam tweet terkait pemindahan ibu kota negara ke Nusantara. Kata "pemindahan" adalah yang paling menonjol, menunjukkan bahwa topik ini adalah yang paling banyak dibahas. Selain itu, kata "negara" dan "pembangunan" juga sering muncul, yang menandakan bahwa banyak diskusi berfokus pada dampak nasional dan pembangunan yang terkait dengan pemindahan ibu kota. Kata "presiden" dan "Jokowi" menunjukkan bahwa peran Presiden Joko Widodo (Jokowi) dalam keputusan pemindahan ibu kota menjadi sorotan utama. Kata "Kalimantan" mengacu pada lokasi baru ibu kota di Pulau Kalimantan. Kata-kata lain seperti "menteri", "pemerintah", "timur", dan "daerah" juga muncul, menunjukkan berbagai aspek dari pemindahan ibu kota yang dibahas, termasuk peran pemerintah, lokasi geografis, dan dampak regional. Kesimpulannya, wordcloud ini memberikan gambaran visual mengenai topik dan tema utama yang muncul dalam diskusi tentang pemindahan ibu kota ke Nusantara di media sosial Twitter.

3.1.6 Deployment

Deployment adalah tahap terakhir dalam menyusun laporan hasil dari kegiatan data mining. Laporan akhir ini mencakup pengetahuan yang berhasil diperoleh serta pengenalan pola yang terjadi dalam proses data mining.

3.2 Pembahasan

Penelitian ini menggunakan algoritma *Naive Bayes* untuk menganalisis sentimen publik terhadap rencana pemindahan ibu kota Nusantara. Algoritma ini terbukti mampu mengklasifikasikan data dengan baik, seperti yang juga ditunjukkan oleh penelitian-penelitian sebelumnya. Sebagai contoh, penelitian oleh Ainurrohma (2021) menunjukkan bahwa penggunaan *RapidMiner* dan *Weka* dapat meningkatkan akurasi klasifikasi dalam analisis sentimen, mirip dengan hasil yang ditemukan dalam penelitian ini di mana model *Naive Bayes* berhasil mencapai akurasi sebesar 92,82%. Hal ini menunjukkan bahwa algoritma ini efektif dalam memproses data berukuran besar yang diambil dari media sosial, seperti Twitter.

Hasil yang diperoleh dari penelitian ini konsisten dengan penelitian oleh Darwis *et al.* (2021), yang menunjukkan bahwa algoritma *Naive Bayes* memiliki kemampuan yang baik dalam mengklasifikasikan sentimen pada data Twitter, khususnya dalam data publik yang berhubungan dengan kebijakan nasional. Dalam penelitian Darwis, sentimen terkait layanan BMKG di Twitter berhasil diidentifikasi dengan akurasi yang tinggi, sejalan dengan temuan penelitian ini mengenai sentimen publik terhadap ibu kota Nusantara. Proses *data preparation* yang melibatkan tokenisasi, normalisasi, dan penghapusan *stop words* menjadi tahap yang sangat penting dalam memastikan kualitas data yang digunakan. Hal ini selaras dengan temuan Permadi (2020) dalam analisis sentimen ulasan

restoran di Singapura, di mana pembersihan data yang baik meningkatkan hasil klasifikasi secara signifikan. Oleh karena itu, proses *cleansing* dan penghapusan elemen yang tidak relevan sangat penting untuk menghasilkan data yang siap untuk dianalisis, meningkatkan keakuratan model. Dalam penelitian ini, *Naïve Bayes* menunjukkan performa yang lebih baik dalam mendeteksi sentimen positif dibandingkan dengan sentimen negatif, dengan presisi untuk sentimen positif mencapai 94,89%.

Hasil ini sejalan dengan penelitian oleh Kusuma dan Nugroho (2021), yang juga menemukan bahwa algoritma ini lebih unggul dalam mendeteksi pola sentimen yang lebih mudah dikenali, terutama sentimen positif. Sebaliknya, kemampuan *Naïve Bayes* dalam mengidentifikasi sentimen negatif masih memerlukan peningkatan, seperti yang juga diamati oleh Juanita (2020) dalam analisis sentimen Pemilu 2019 di Twitter, di mana algoritma lebih sensitif terhadap sentimen positif daripada negatif. Selain itu, penelitian oleh Hasri dan Alita (2022) menunjukkan bahwa kombinasi antara *Naïve Bayes* dan *Support Vector Machine* dapat menghasilkan klasifikasi yang lebih akurat dalam analisis sentimen terkait dampak pandemi COVID-19 di Twitter. Namun, dalam penelitian ini, penggunaan algoritma *Naïve Bayes* secara tunggal sudah memberikan hasil yang cukup memuaskan dengan akurasi yang tinggi. Meski demikian, peningkatan pada tahap evaluasi dan penerapan metode tambahan seperti yang disarankan oleh Hasri dan Alita dapat memberikan hasil yang lebih optimal.

4. Kesimpulan

Kesimpulan pada penelitian ini menunjukan bahwa pandangan masyarakat Indonesia terhadap rencana pemindahan ibu kota memiliki banyak pandangan. Oleh karena itu penelitian ini bermaksud untuk mengetahui pandangan masyarakat Indonesia terhadap rencana pemindahan Ibu Kota Negara pada media sosial twitter menggunakan pendekatan data mining yaitu metode algoritma *Naïve Bayes*. Serta dalam penelitian ini menggunakan mode pengembangan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Hasil dari proses pengujian mendapatkan nilai akurasi 92.82%, model mengklasifikasikan sekitar 93% *tweet* dengan benar. Secara keseluruhan, model ini efektif dalam mengklasifikasikan sentimen.

5. Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan bantuan selama proses pembuatan jurnal ini.

6. Daftar Pustaka

- Ainurrohmah, A. (2021, February). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 493-499).
- Ananda, D., & Suryono, R. R. (2024). Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan *Naïve Bayes*. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(2), 748-757. DOI: <http://dx.doi.org/10.30865/mib.v8i2.7517>.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.

- Hakim, Z. R., & Sugiyono, S. (2024). Analisa Sentimen Terhadap Kereta Cepat Jakarta–Bandung Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor. *Jurnal Sains dan Teknologi*, 5(3), 939-945. DOI: <https://doi.org/10.55338/saintek.v5i3.1423>.
- Hasri, C. F., & Alita, D. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal informatika dan rekayasa perangkat lunak*, 3(2), 145-160. DOI: <https://doi.org/10.33365/jtk.v15i1.744>.
- Juanita, S. (2020). Analisis sentimen persepsi masyarakat terhadap pemilu 2019 pada media sosial twitter menggunakan naive bayes. *Jurnal Media Informatika Budidarma*, 4(3), 552-558. DOI: <http://dx.doi.org/10.30865/mib.v4i3.2140>.
- Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(2), 137-146. DOI: <https://doi.org/10.32815/jitika.v15i2.557>.
- Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 222-238. DOI: <https://doi.org/10.32520/stmsi.v10i1.1179>.
- Permadi, V. A. (2020). Analisis sentimen menggunakan algoritma Naïve Bayes terhadap review restoran di Singapura. *Jurnal Buana Informatika*, 11(2), 141-151. DOI: <https://doi.org/10.24002/jbi.v11i2.3769>.
- Safira, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONAsi: Jurnal Sistem Informasi*, 5(1), 59-70. DOI: <https://doi.org/10.31849/zn.v5i1.12856>.