

PENERAPAN METODE MACHINE LEARNING - NAIVE BAYES PADA ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA BARU

Implementation of Machine Learning – Naïve Bayes Method to Sentiment Analysis of New Country Capitals

David J. Putnarubun¹, C. F. Palembang²

^{1,2}Program Studi Ilmu Komputer, FMIPA Universitas Pattimura

e-mail: ¹daudputnarubun@gmail.com

Abstrak

Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap pemindahan ibu kota negara (IKN) pada twitter di tahun 2023 apakah masih menjadi suatu kontroversi atau sudah lebih beraroma positif dibandingkan dengan opini masyarakat pada saat pertama kali isu ini naik ke media sosial di tahun 2019. Penelitian ini menggunakan metode naïve bayes dalam melakukan klasifikasi data, diharapkan dapat menyajikan hasil yang akurat. Hasil yang diperoleh dari analisis sentimen terhadap pemindahan ibu kota negara (ikn) baru dengan presentase nilai positif sebesar 55,85% dan sentimen negatif sebesar 45,15. Itu artinya respon dari masyarakat terhadap isu IKN di media sosial pada tahun 2023 sudah lebih banyak positif menerima pemindahan ibu kota negara dengan segala urgensinya dibandingkan dengan pada tahun 2019 saat Presiden Joko Widodo pertama kali mengumumkan wacana tersebut. Diperoleh nilai akurasi (accuracy) sebesar 99,12%, nilai akurasi recall untuk hasil negatif yaitu 98,37% dan hasil positif 99,71%. Kemudian untuk nilai akurasi precision untuk pred negatif yaitu 99,63% dan pred positif 99,72%. Dengan demikian maka metode naïve bayes memiliki nilai akurasi yang cukup tinggi sehingga dapat dijadikan pilihan metode pada penelitian selanjutnya.

Kata Kunci: IKN, Analisis Sentimen, Naïve Bayes, Machine Learning, Twitter

Abstract

This study aims to analyze public sentiment towards the relocation of the national capital (IKN) on Twitter in 2023 whether it is still a controversy or is it already more positive compared to public opinion when this issue first surfaced on social media in 2019. This research uses the naïve Bayes method in classifying data is expected to present accurate results. The results obtained from the sentiment analysis for the relocation of the new national capital (ikn) with a positive percentage value of 55.85% and a negative sentiment of 45.15. This means that the response from the public to the IKN issue on social media in 2023 is better to accept the relocation of the national capital with all its urgency compared to 2019 when President Joko Widodo first announced the discourse. The accuracy value (accuracy) is 99.12%, the recall accuracy value for negative results is 98.37% and positive results is 99.71%. Then for the precision accuracy value for negative pred is 99.63% and positive pred is 99.72%. Thus, the naïve Bayes method has a fairly high accuracy value so that it can be used as a method of choice for further research.

Keywords: IKN, Sentiment Analysis, Naïve Bayes, Machine Learning, Twitter



<https://doi.org/10.30598/parameter2i01pp57-66>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

1. PENDAHULUAN

Pada tahun 2019, Presiden Joko Widodo mengumumkan pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur yang dinamakan dengan Nusantara dengan berbagai macam kajian dan konsultasi publik, kemudian dicantumkan dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) Tahun Anggaran 2020-2024 [1][2][3]. Beberapa keharusan mendesak yang menjadi alasan pemindahan ibu kota negara oleh Presiden Joko Widodo antara lain; (1) Kepadatan penduduk di pulau Jawa yang sebesar 56,56 persen menurut Survei Penduduk Antar Sensus (SUPAS) tahun 2015, (2) Kontribusi ekonomi pulau Jawa terhadap pertumbuhan ekonomi Indonesia (PDB) yang sangat mendominasi sebesar 58,49 persen, (3) Ketersediaan air bersih yang sudah mengalami krisis, (4) Proporsi konsumsi lahan terbangun di pulau Jawa mendominasi berdasarkan hasil modelling KLHS Bappenas 2019 mencapai lima kali lipat dari Kalimantan[2].

Keputusan pemerintah memindahkan ibu kota negara menimbulkan berbagai opini di masyarakat diantaranya melalui media sosial (*twitter*). *Twitter* merupakan salah satu media sosial dan layanan *microblogging* yang memungkinkan mengirimkan pesan (*tweet*) secara *realtime* oleh penggunanya. *Tweet* dibatasi hanya sampai 140 karakter sehingga seringkali mengandung singkatan dan bahasa gaul (*slang*)[4]. Setiap hari, ribuan bahkan jutaan *tweet* opini muncul terhadap berbagai isu yang sedang terjadi seperti pendapat atau ekspresi kepada suatu jasa pelayanan, fasilitas maupun isu ekonomi, sosial maupun politik yang sedang trend diperbincangkan akhir-akhir ini[3] termasuk pemindahan ibu kota negara ke Kalimantan.

Berbagai *tweet* opini pro dan kontra masyarakat dengan perspektif yang berbeda terkait pemindahan ibu kota negara pernah menjadi trending topik pada tahun 2019 dan menyebabkan perdebatan antar pengguna, dan pada tanggal 31 Mei tahun 2023 saat Presiden Joko Widodo mengumumkan Logo IKN terpilih, isu pemindahan ibu kota negara kembali hangat diperdebatkan tentunya dengan kata-kata atau kalimat yang bernada positif maupun negatif.

Kecenderungan pengguna *twitter* memposting pesan atau konten dapat diketahui dengan cara analisis sentimen[5]. Teori analisis sentimen bertujuan untuk mengelompokkan polaritas pada sebuah teks yang ada di dalam kalimat, fitur/aspek, dan dokumen yang dapat ditetapkan memiliki sentimen positif atau negatif[6].

Penelitian di bidang *text mining* dengan teknik analisis sentimen semakin berkembang terlihat pada berbagai publikasi jurnal ilmiah, menggunakan berbagai macam metoda atau algoritma yang telah dilakukan oleh peneliti sebelumnya.

Pada penelitian sebelumnya, analisis sentimen menggunakan metode *naïve bayes* terhadap program kampus merdeka menghasilkan klasifikasi sentimen positif sebanyak 272 opini dan 229 sentimen negatif dengan rata-rata akurasi 60%, presisi 64%, recall 58%, dan *f1-score* sebesar 58%[7].

Kemudian penelitian lain menggunakan metode yang sama *naïve bayes* pada analisis sentimen mengenai vaksin covid-19 menghasilkan persentase akurasi sebesar 100%[8].

Penelitian analisis sentimen lainnya terhadap pembelajaran daring dengan metode *naïve bayes* menghasilkan 30% sentimen positif, 69% sentimen negatif dan 1% netral [9].

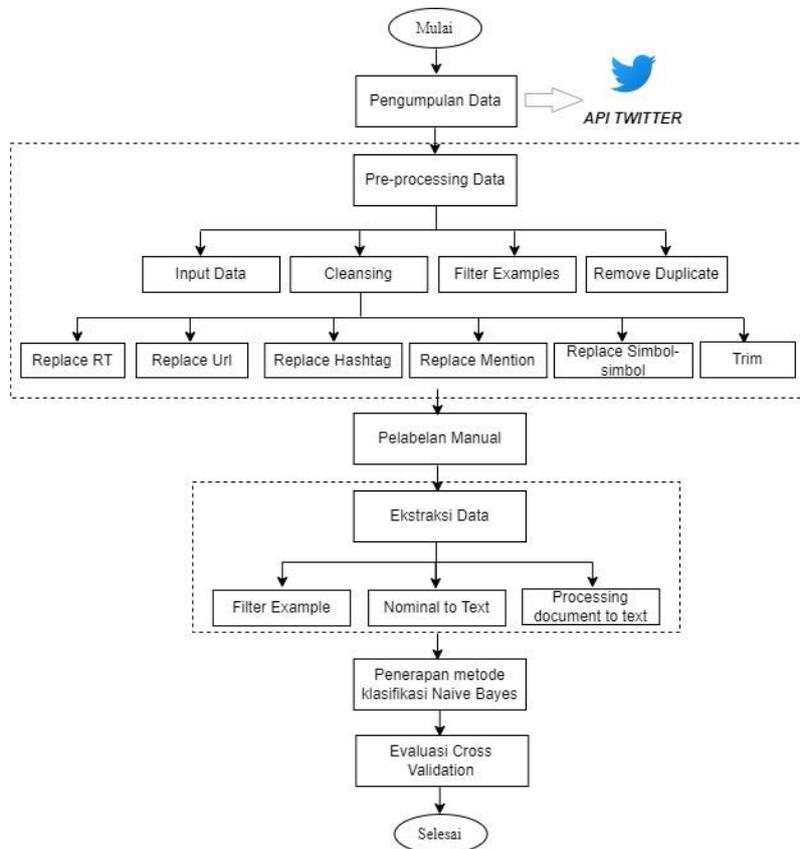
Masih dengan tema penelitian yang hampir sama terkait analisis sentimen terhadap pembelajaran campuran menggunakan *naïve bayes* menghasilkan polarisasi sentimen positif dan negatif hampir seimbang yaitu 44,51% positif dan 45,80% negatif [10].

Berikutnya dengan topik yang sama mengenai analisis sentimen terhadap ibu kota negara dan metode yang sama naïve bayes dengan penelitian saat ini, sebelumnya di tahun 2019 Natasuwarna dalam penelitiannya menyatakan menghasilkan nilai akurasi rata-rata sebesar 89,86% dari 200 data twitter diperoleh 100 opini positif dan 100 opini negatif dengan menggunakan lima rasio perbandingan data latih dan data uji[11].

Dari beberapa hasil penelitian yang telah dilakukan sebelumnya, penelitian ini bertujuan menganalisis sentimen masyarakat terhadap pemindahan ibu kota negara (IKN) pada twitter di tahun 2023 apakah masih menjadi suatu kontroversi atau sudah lebih beraroma positif dibandingkan dengan opini masyarakat pada saat pertama kali isu ini naik ke media sosial di tahun 2019. Penelitian ini menggunakan metode naïve bayes dalam melakukan klasifikasi data, diharapkan dapat menyajikan hasil yang akurat.

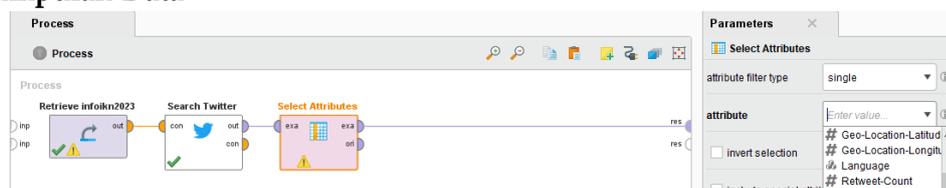
2. METODE PENELITIAN

Pada bagian metodologi ini membahas tentang tahapan penelitian yang dilakukan dalam proses analisis sentimen terhadap data twitter, sehingga menjadi informasi pengklasifikasian. Berikut tahapan yang digambarkan pada Gambar 1.



Gambar 1. Tahap penelitian analisis sentimen

a. Pengumpulan Data

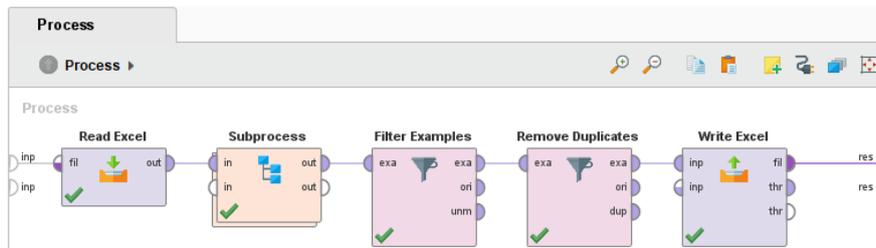


Gambar 2. Proses pengumpulan data twitter menggunakan RapidMiner

Proses pengumpulan data menggunakan teknik *crawling data*, yaitu data diperoleh dari media sosial *twitter*, prosesnya dilakukan secara manual melalui aplikasi RapidMiner dengan mengakses token API *twitter*.

b. *Pre-processing Data*

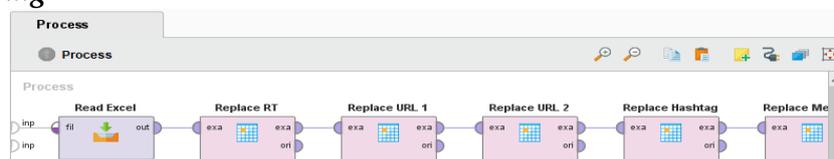
Pada proses *pre-processing*, data mentah disiapkan menjadi data set melalui proses pengolahan data untuk digunakan sebagai model dalam pembelajaran mesin[12]. Adapun tahapan dari *pre-processing* data yaitu:



Gambar 3. Tahapan *pre-processing* data

Tahapan *pre-processing* data yaitu dimulai dari proses membaca *file excel* yang sebelumnya sudah disimpan ketika *dicrawling*, kemudian lakukan pembersihan data (*cleansing*) yang disimpan di dalam satu lokasi khusus dinamakan *subprocess*, kemudian dilanjutkan dengan menggunakan operator *filter examples* untuk menghilangkan atau menghapus kolom data *tweet* yang kosong, selain itu gunakan juga operator *remove duplicates* untuk menghapus data ganda atau data *tweet* yang sama atau dicetak berulang. Setelah semua proses selesai, data disimpan ke dalam *file excel* yang baru sebagai dataset untuk proses analisis sentimen.

2.2.1 *Cleansing*



Gambar 4. Tahapan *cleansing* data

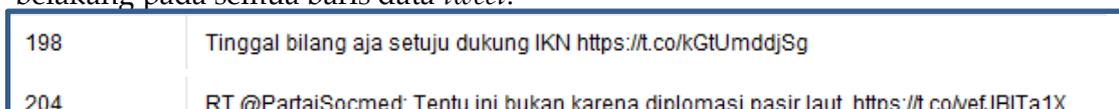
Pada tahap *cleansing* data, terdapat beberapa atribut yang digunakan, yaitu:

1. *Replace RT*, artinya proses menghapus semua kata RT (*retweet*) pada semua baris *tweet* yang mengandung karakter RT.



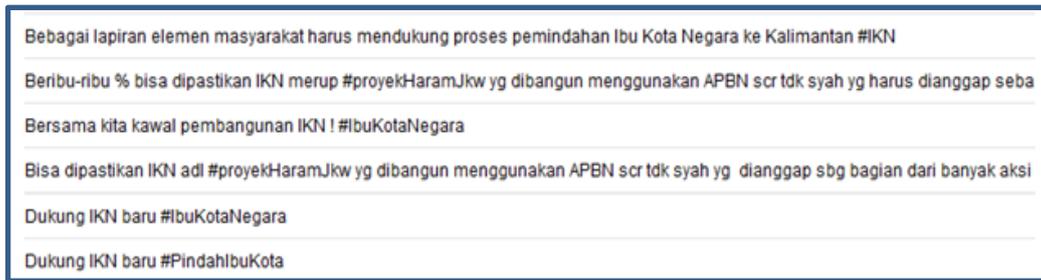
Gambar 5. kalimat mengandung RT (*retweet*)

2. *Replace URL*, artinya proses menghapus semua *url/link*, baik itu di depan, tengah atau belakang pada semua baris data *tweet*.



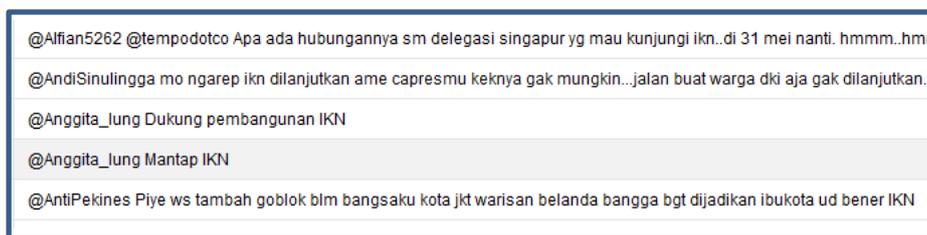
Gambar 6. kalimat mengandung url

3. *Replace Hashtag*, artinya proses menghapus semua karakter *hashtag* (#), baik itu di depan, tengah atau belakang pada semua baris data *tweet*.



Gambar 7. kalimat mengandung *Hashtag* (#)

4. *Replace Mention*, artinya proses menghapus semua karakter *mention* (@), baik itu di depan, tengah atau belakang *tweet* pada semua baris data *tweet*.



Gambar 8. kalimat mengandung *Mention* (@)

5. *Replace Simbol*, artinya proses menghapus semua karakter simbol (!\$%^&*;,:'"), pada semua baris data *tweet*.



Gambar 9. kalimat mengandung karakter simbol (!\$%^&*;,:'")

6. *Trim*, gunanya untuk menghapus *white spaces* (spasi-spasi yang tidak diperlukan) pada semua baris data *tweet*.

2.2.2 Filter Examples

Pada tahap *pre-processing* menggunakan operator *filter examples* gunanya untuk menghapus baris data yang kosong (hanya berisikan tanda tanya).

1836	????????..'
1869	????????..'
1895	????????..'

Gambar 10. Baris data *tweet* kosong

2.2.3 Remove Duplicate

Operator *remove duplicate* gunanya untuk menghapus baris data *tweet* ganda atau yang berulang.

3468	Sektor UMKM mengalami pertumbuhan pesat sejak pemerintah memulai pembangunan infrastruktur IKN. #IKNPemerataanEkonomi #IKNS...
2678	Sektor UMKM mengalami pertumbuhan pesat sejak pemerintah memulai pembangunan infrastruktur IKN. #IKNPemerataanEkonomi #IKNS...
3471	Sektor UMKM mengalami pertumbuhan pesat sejak pemerintah memulai pembangunan infrastruktur IKN. #IKNPemerataanEkonomi #IKNS...
2680	Sektor UMKM mengalami pertumbuhan pesat sejak pemerintah memulai pembangunan infrastruktur IKN. #IKNPemerataanEkonomi #IKNS...

Gambar 11. Baris data *tweet duplicate*

c. **Pelabelan Manual**

Pemberian label sentimen positif dan negatif akan diberikan setelah proses *cleaning* selesai, dan datanya akan dijadikan data *training* atau data latih sebagai pembelajaran untuk metode *naïve bayes*.

d. **Ekstrasi Data**

Ekstrasi data/fitur merupakan tahap perhitungan kata, perhitungan vector dan TF-IDF. Pada proses ini akan menghasilkan model yang akan digunakan untuk ketepatan hasil klasifikasi. Terdiri dari proses filter examples, kemudian nominal to text untuk dapat melakukan proses processing to text.

e. **Penerapan Metode Klasifikasi *Naïve Bayes***

Setelah diset sentimen terhadap data latih (*training*) sebanyak yang diinginkan untuk dipelajari oleh *naïve bayes*, kemudian dapat dilakukan proses klasifikasi sentimen terhadap data lainnya yang belum dilabeli secara otomatis, kemudian lakukan penyimpanan model.

f. **Evaluasi *Cross Validation***

Tahap terakhir analisis sentimen pada penelitian ini yaitu evaluasi *cross validation*, untuk menilai keakuratan sebuah model.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Berdasarkan hasil *crawling* data *twitter* dengan *query* *ikn* limit data yang peneliti tentukan sebanyak 5.000 data dan batasan hanya *tweet* yang berbahasa indonesia, hasilnya data yang diperoleh sebanyak 3.740, yang kemudian disimpan ke dalam format *xlsx*.

Row No.	Id	Text
1	1663724943...	Selamat pagi. Inilah logo Ibu Kota Nusantara yang diluncurkan kemarin. Logo yang akan menjadi identitas visual bagi IK...
2	1663669377...	Keputusan Presiden Jokowi membuka tambang dan ekspor pasir laut, semakin banyak yang mengkritisi. Termasuk gur...
3	1664448592...	Mengerikan sekali besarnya utang Indonesia yang tembus Rp20750 triliun. Kata Pak JK berutang itu mudah, yang susa...
4	1666066335...	RT @panglima_z: Janji Pilpres :
5	1666066025...	RT @aik_arif: Calon IKN... Laju Kenaikan Suhu di Kalimantan Timur Tertinggi di Indonesia - https://t.co/gzgaBZFfFB https...
6	1666065988...	RT @Hudakey: Saya sedang membayangkan kira" apa jawabannya Tentang Kondisi Jakarta Pasca Pindah ke IKN, Jika...
7	1666065950...	RT @aik_arif: Calon IKN... Laju Kenaikan Suhu di Kalimantan Timur Tertinggi di Indonesia - https://t.co/gzgaBZFfFB https...
8	1666065815...	RT @Hudakey: Saya sedang membayangkan kira" apa jawabannya Tentang Kondisi Jakarta Pasca Pindah ke IKN, Jika...
9	1666065790...	RT @Hudakey: Saya sedang membayangkan kira" apa jawabannya Tentang Kondisi Jakarta Pasca Pindah ke IKN, Jika...
10	1666065733...	RT @KurawalFound_ID: PETAKA LINGKUNGAN PRESIDEN JOKOWI: BEDA UCAPAN & KENYATAAN
11	1666065602...	RT @Hudakey: Saya sedang membayangkan kira" apa jawabannya Tentang Kondisi Jakarta Pasca Pindah ke IKN, Jika...
12	1666065539...	RT @aik_arif: Calon IKN... Laju Kenaikan Suhu di Kalimantan Timur Tertinggi di Indonesia - https://t.co/gzgaBZFfFB https...
13	1666065260...	RT @Hudakey: Saya sedang membayangkan kira" apa jawabannya Tentang Kondisi Jakarta Pasca Pindah ke IKN, Jika...
14	1666065107...	RT @__wantaNKRI: Bacapres PDIP @ganjarpranowo akan lanjutkan proyek IKN dan proyek ini menjadi target utama di...
15	1666064953...	RT @KurawalFound_ID: PETAKA LINGKUNGAN PRESIDEN JOKOWI: BEDA UCAPAN & KENYATAAN

ExampleSet (3,740 examples, 1 special attribute, 1 regular attribute)

Gambar 12. Hasil pengumpulan data twitter

3.2 Pre-processing data

Setelah melakukan proses pembersihan data yang dimulai dari menghapus kata

dan karakter yang tidak dibutuhkan serta kolom baris yang tidak ada isinya pada tahap *pre-processing* data, diperoleh dataset yang digunakan pada tahap berikutnya yaitu sebanyak 1.564 data.

Row No.	Text
869	ni apa sih tujuannya? biar Jakarta beneran tenggelem? trs biar pd pindah ke IKN? biar berhasil proyeknya? gila ya
870	Ada apa ini pak
871	Untuk biaya pemilu atau untu biaya IKN pak?
872	Jadi ingat IKN...sampai 190 tahun ..dll
873	terus di IKN kualitas udara ngedrop setelah warga makin padat... jadi lingkaran setan...
874	Dukung Penuh IKN Nusantara untuk Indonesia lebih baik
875	Dukung terus pembangunan IKN Nusantara di Kaltim
876	Dukung Pembangunan IKN yang berdampak baik kemajuan bangsa Indonesia dan Ekonomi Masyarakat
877	Emang omongannya presidennya gitu kan?
878	Waktu masih jadi gubernur ngomongnya ga bakalan nyapres.. Eh baru 2 tahun udh ikutan nyapres..
879	Waktu ngomongin masalah oversubscribe investor IKN... Nyatanya Subscriber abal abal..
880	Saat ini progres pembangunan istana presiden di IKN Nusantara sudah mencapai 15%. Pembangunan Istana ini ditargetkan dap
881	Kan Ada 11 ribu triliun dikantono kok Gak diokake banoun IKN?

ExampleSet (1,564 examples, 0 special attributes, 1 regular attribute)

Gambar 13. Dataset setelah dilakukan *pre-processing*

3.3 Pelabelan Manual

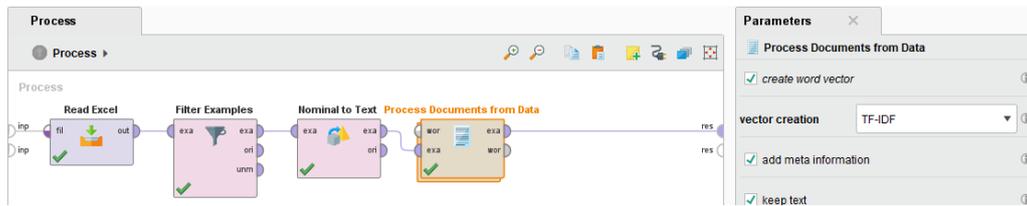
Dari 1.564 dataset, sebanyak 313 data dilabeli sentimen secara manual untuk dijadikan sebagai data latih (*training*). Label klasifikasi dikategorikan menjadi dua, yaitu kategori sentimen positif dan sentimen negatif. Rasio yang digunakan adalah 20% : 80% data latih (*training*) dan data uji (*testing*).

183	AhSial Ihu bawa2 Si Bokep ke IKN	Negatif
184	GEROMBOLAN Sebelah Pasti Sedang Caricari Narasi Tentang Logo IKN Di Kaitkan Dengan Sesuatu yg Negatif	Negatif
185	Kalo gak mangkrak ya salah posisinya Akhirnya bisa mangkrak juga Insfratruktur tanpa perencanaan yg matan...	Negatif
186	Banyak Hutan dibabat terus logo ikn pohon hayat Ga sinkron Sama seperti dulu ucapan g sesuai dengan kenyataan Ga impor ternyata	Negatif
187	Logo IKN apa BIN njir haha	Negatif
188	Keren euy januari udah banyak berdiri bangunan di ikn	Positif
189	Ikn ini paling paling bernasib sama dg Esemka terdengar tapi tak terlihat	Negatif
190	Kalo IKN uda jadi 100 pengen main kesana lihat seberapa wow nya ibukota baru	Positif
191	Dan semoga Jekardah jd sepi ga macet bebas polusi hihi	Positif
192	Tak ada yang melarang Kritik sepanjang itu disampaikan dgn Data dan kalkulasi yg jelas alias bukan opini buta IKN harus berlanjut krn	Positif
193	Ibu Kota yg ada saat ini adlh Warisan Belanda kita harus bermartabat Soal Hubungan Kerjasama yg berdampak TKA China tanya Paloh	Positif

Gambar 14. Data latih (*training*) yang dilabeli sentimen secara manual

3.4 Ekstraksi Data (Pemrosesan TF-IDF)

Algoritma TF-IDF (*Term-Frequency-Inverse Document*) berfungsi untuk menghitung bobot dari setiap kata. Proses TF-IDF meliputi *tokenize*, *transform cases*, *filterstopwords* dan *filter tokens*) yang disimpan dalam satu subfolder *Process Documents form Data*, dimana proses tersebut dikoneksikan dengan dataset yang telah *dipre-processing* sebelumnya.



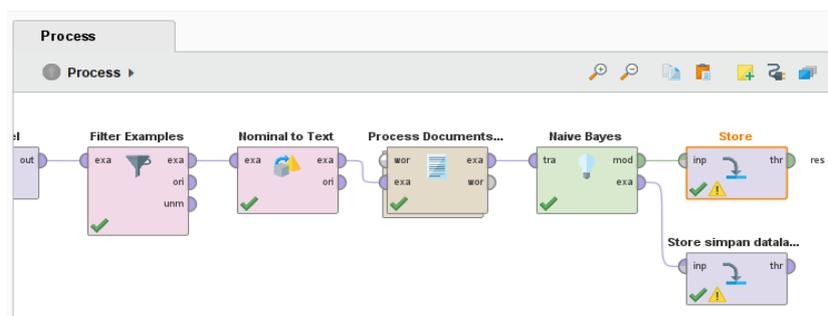
Gambar 15. Proses TF-IDF

	yan	yayaan	yeah	yess	yg	yhaaa	yntkts	yo	yogyakart	yokk	york	youtube	yth	yuk	yulis	zaman	zero	zonk	text	
2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
10	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
11	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
12	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
13	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
14	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
15	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
16	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
17	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
18	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
19	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
20	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
21	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
22	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
23	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
24	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Gambar 16. Data hasil proses TF-IDF

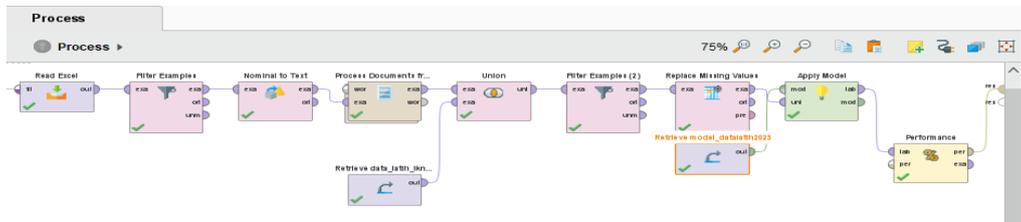
3.5 Implementasi Metode Klasifikasi Naïve Bayes

Proses berikutnya menerapkan metode klasifikasi naïve bayes. Naïve bayes mempelajari pola data kemudaiian menghasilkan model klasifikasi untuk dipakai dalam menentukan analisis sentimen dari *tweet-tweet* lain yang belum dilabeli analisis sentimen.



Gambar 17. Proses pembuatan model klasifikasi

Setelah dibentuk model klasifikasi menggunakan metode naïve bayes, kemudian atributnya akan digabungkan dengan data uji (*testing*) karena kata-kata yang terdapat pada data latih (*testing*) berbeda dengan yang ada pada data uji (*testing*). Selanjutnya dilakukan proses pengujian performa dataset yang akan menghasilkan nilai *Accuracy*, *Recall*, dan *Precision*.



Gambar 18. Proses pengujian perfomansi dataset

3.6 Evaluasi Cross Validation

Pada tahap akhir evaluasi cross validation meliputi proses latih (*training*) dan uji (*testing*) (*apply model & performance*) menggunakan metode naïve bayes untuk memperoleh nilai *accuracy*, *precision*, dan *recall* berdasarkan *performance vector*.

Tujuan tahap evaluasi yaitu menentukan nilai dari kegunaan model yang telah berhasil dibuat pada tahapan sebelumnya, menggunakan *Confusion Matrix* : *true positive rate* (TP rate), *true negative rate* (TN rate), *false positive rate* (FP rate) dan *false negative rate* (FN rate) sebagai indikator [13].

accuracy: 99.12%

	true Negatif	true Positif	class precision
pred. Negatif	544	2	99.63%
pred. Positif	9	696	98.72%
class recall	98.37%	99.71%	

Gambar 19. Nilai Akurasi (Accuracy)

Nilai akurasi yang diperoleh adalah sebesar 99,12%, nilai akurasi *recall* untuk hasil negatif yaitu 98,37% dan hasil positif 99,71%. Kemudian untuk nilai akurasi *precision* untuk pred negatif yaitu 99,63% dan pred positif 99,72%.

PerformanceVector

```
PerformanceVector:
accuracy: 99.12%
ConfusionMatrix:
True:  Negatif Positif
Negatif:  544    2
Positif:   9    696
```

Gambar 20. Performance Vector

Berdasarkan proses latih (*training*) dan pengujian (*testing*), maka diperoleh hasil:

Tabel 1. Data latih (*training*) dan data uji (*testing*)

Klasifikasi	Data latih (<i>training</i>)	Data uji (<i>testing</i>)	Total
Positif	160	698	858
Negatif	153	553	706
Total	313	1251	1564

4. KESIMPULAN

Berdasarkan pengujian terhadap data *tweet* dengan *query* pencarian yaitu IKN (Ibu Kota Negara) Baru, maka dapat disimpulkan bahwa:

1. Hasil klasifikasi sentimen positif sebesar 55,85% dan sentimen negatif sebesar 45,15%.

Hal ini menunjukkan bahwa respon dari masyarakat terhadap isu IKN di media sosial pada tahun 2023 sudah lebih banyak positif menerima pemindahan ibu kota negara dengan segala urgensinya dibandingkan dengan pada tahun 2019 saat Presiden Joko Widodo pertama kali mengumumkan wacana tersebut.

2. Diperoleh nilai akurasi (*accuracy*) sebesar 99,12%, nilai akurasi *recall* untuk hasil negatif yaitu 98,37% dan hasil positif 99,71%. Kemudian untuk nilai akurasi *precision* untuk pred negatif yaitu 99,63% dan pred positif 99,72%. Dengan demikian maka metode naïve bayes memiliki nilai akurasi yang cukup tinggi sehingga dapat dijadikan pilihan metode pada penelitian selanjutnya

DAFTAR PUSTAKA

- [1] F. Hadi and R. Ristawati, "Pemindahan Ibu Kota Indonesia dan Kekuasaan Presiden dalam Perspektif Konstitusi," *J. Konstitusi*, vol. 17, no. 3, pp. 530–557, 2020, doi: 10.31078/jk1734.
- [2] D. S. Wulandari, A. F. Ariza, and N. Nurhayati, "Pengaruh Pemindahan Ibu Kota Negara terhadap Investasi di Kabupaten Tabalong Kalimantan Selatan," *J. Educ.*, vol. 5, no. 2, pp. 3804–3807, 2023, doi: 10.31004/joe.v5i2.1063.
- [3] R. K. Septiani, S. Anggraeni, and S. D. Saraswati, "Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *Teknika*, vol. 16, no. 02, pp. 1–5, 2022.
- [4] H. Nurrin Muchammad Shiddieqy, S. Paulus Insap, and W. Wing Wahyu, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. March, pp. 57–64, 2016.
- [5] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naive Bayes Classifier," *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.
- [6] S. Lestari, M. Mupaat, and A. Erfina, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter," *JUSIFO (Jurnal Sist. Informasi)*, vol. 8, no. 1, pp. 13–22, 2022, doi: 10.19109/jusifo.v8i1.12116.
- [7] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *J. Sist. Komput. dan Inform. Hal 296–*, vol. 301, no. 2, pp. 25–38, 2022.
- [8] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naive Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [9] A. Tanggu Mara, E. Sedyono, and H. Purnomo, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 24–31, 2021, doi: 10.53682/jointer.v2i01.30.
- [10] R. Watrionthos, M. Giatman, W. Simatupang, R. Syafriyeti, and N. K. Daulay, "Analisis Sentimen Pembelajaran Campuran Menggunakan Twitter Data," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 166, 2022, doi: 10.30865/mib.v6i1.3383.
- [11] A. P. Natasuwarna, "Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes," *Sensitif (Seminar Nas. Sist. Inf. dan Tek. Inform.)*, pp. 47–53, 2019.
- [12] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [13] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 2, no. 1, pp. 32–41, 2017, doi: 10.31284/j.integer.2017.v2i1.95.