

Penerapan Algoritma Naïve Bayes Untuk Analisis Pandangan Masyarakat Terhadap Pemindahan Ibu Kota Indonesia

Aditiya Ihzar Eka Prayogo¹, Painem²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

^{1,2}Jakarta Selatan, DKI Jakarta, Indonesia

1811500584@student.budiluhur.ac.id¹, painem@budiluhur.ac.id²

Abstrak. Pemerintah sudah lama mengumumkan rencana untuk memindahkan ibu kota negara, tetapi masih terdapat kontroversi di antara masyarakat yang setuju dan tidak setuju dengan rencana tersebut. Masyarakat memiliki alasan dan pandangan yang berbeda-beda tentang rencana ini. Untuk mendapatkan gambaran umum tentang respon masyarakat terhadap rencana pemindahan ibu kota, perlu melihat berbagai reaksi yang berbeda. Dengan era teknologi yang sangat canggih seperti sekarang ini, membuat penilaian sentimen publik terhadap berbagai isu menjadi mudah dan cepat. Hal ini disebabkan oleh maraknya penggunaan media sosial oleh masyarakat. Media sosial saat ini dapat digunakan sebagai media hiburan, tetapi juga sebagai media untuk menyatakan pendapat dan pandangan tentang berbagai topik. Begitu juga dengan topik pemindahan ibu kota negara, dimana masyarakat banyak mengekspresikan reaksinya melalui berbagai media, termasuk media sosial seperti Twitter. Twitter adalah salah satu media sosial yang banyak digunakan oleh masyarakat Indonesia untuk menyatakan pendapatnya. Pendapat tentang pemindahan ibu kota yang diperoleh dari Twitter perlu dianalisis. Pada penelitian ini, metode analisis sentimen yang digunakan untuk menganalisa topik pemindahan ibu kota dari media sosial twitter adalah metode Naive Bayes Classifier (NBC). Hasil analisis berdasarkan 1.272 tweet menunjukkan bahwa sentimen positif sebesar 86.95 % dan sentimen negatif sebesar 13.05 % pada periode 10 May 2022 hingga 1 Juni 2022, sementara hasil pengujian yang diperoleh adalah nilai akurasi sebesar 93 %, presisi sebesar 87 % dan recall sebesar 100 %.

Kata Kunci: analisis sentimen, twitter, klasifikasi, naïve bayes

1 Pendahuluan

Rencana pemindahan ibu kota ke Kalimantan Timur secara resmi di deklarasikan pada bulan agustus 2019 oleh Presiden Joko Widodo. Namun dibalik hal tersebut terdapat kontroversi dari masyarakat yang merasa pro dan kontra terhadap pemindahan ibu kota. Saat ini, masyarakat dalam mengemukakan pendapat (komentar) atau dalam melakukan penilaian bisa dilakukan melalui media Sosial. Salah satu media sosial yang digunakan dalam memberikan komentar atau penilaian adalah Twitter. Twitter adalah komunitas informal dengan kontribusi mini ke blog yang memiliki jumlah klien dan basis informasi yang besar dan berkembang. Twitter dalam banyak kasus digunakan sebagai tempat untuk menawarkan pendapat karena memungkinkan pengguna untuk berbagi pesan menggunakan pesan singkat yang disebut tweets. Ada sebuah ruang untuk mengarahkan eksplorasi dengan memecah perasaan dan pendapat secara komputasi [1].

Berdasarkan data dari website statista, kemajuan klien Twitter pada tahun 2017 mencapai 109 juta klien dinamis dan pada tahun 2022 mencapai 229 juta klien dinamis di seluruh dunia. Klien web menyusun sentimen dan penilaian tentang berbagai subjek di administrasi Twitter. Saat klien mengekspresikan tema yang berbeda, misalnya, masalah pemerintahan, ini menjadikan Twitter sumber data yang potensial dan mahir. Penelitian ini menggunakan Twitter karena memiliki beberapa manfaat, khususnya digunakan oleh berbagai kelompok klien, memiliki pesan singkat yang disebut tweet yang berisi berbagai kesimpulan umum, terus berkembang, dan menyebarkan berita dengan cepat. Berbagai topik dibicarakan secara luas di Twitter mulai dari masalah keuangan, masalah legislatif, sosial, budaya, regulasi, dan lainnya. Salah satu topik yang pernah trending adalah pemindahan ibu kota ke Kalimantan Timur. Pada periode 10 May 2022 hingga 1 Juni 2022 didapatkan 1.272 tweet yang berhubungan dengan tema pemindahan ibu kota. Berdasarkan 1.272 tweet yang didapatkan maka perlu dilakukan analisa untuk mengetahui pendapat masyarakat tentang pemindahan ibu kota. Baik pendapat masyarakat yang positif maupun yang negatif. Salah satu metode analisis sentimen yang digunakan adalah metode Naive Bayes Classifier (NBC).

Pada penelitian sebelumnya mengenai analisis sentimen yang dilakukan oleh [5] menerapkan metode Naives Bayes Classifier untuk menganalisis sentimen masyarakat Jakarta terhadap moda transportasi baru yaitu MRT. Penelitian menghasilkan akurasi sebesar 75%. Pada penelitian lainnya yang dilakukan oleh [10] juga menggunakan metode KNN untuk menganalisis sentimen masyarakat terhadap pembelajaran dari pada masa

pandemi Covid-19. Fitur yang digunakan adalah CountVectorizer. Hasil akurasi sebesar 76,56%. Metode Naive Bayes merupakan salah satu metode klasifikasi yang populer digunakan dalam data mining dan machine learning. Salah satu kelebihan dari metode ini adalah mudah dipahami dan dipelajari. Metode ini juga memiliki kemampuan untuk menangani data yang berskala besar dengan cepat dan efisien. Selain itu, metode Naive Bayes juga mampu menangani data yang tidak seimbang, dimana data pada kelas tertentu jauh lebih banyak dibandingkan dengan kelas lainnya. Metode ini juga tidak memerlukan banyak data latih, sehingga cocok digunakan pada kasus-kasus dimana jumlah data latih terbatas. Dengan semua kelebihan tersebut, metode Naive Bayes dapat menjadi pilihan yang tepat untuk digunakan dalam berbagai aplikasi klasifikasi data.

Algoritme Naive Bayes untuk mengkonsolidasikan probabilitas sebelumnya dan probabilitas konsistensi dalam persamaan soliter, yang dapat digunakan untuk menghitung probabilitas dari setiap urutan probabilitas yang mungkin. Setelah selesai kami memilih pengaturan dengan nilai terbesar. Dalam penelitian yang diarahkan oleh [3], Peneliti tersebut berhasil mendapatkan akurasi yang sangat akurat dalam melakukan klasifikasi tweet menggunakan Algoritme Naive Bayes yaitu sebesar 99,62%, Naive Bayes juga digunakan oleh [4], untuk mengklasifikasi data tweet yang berjumlah 2000 dan mendapatkan akurasi yang cukup baik yaitu 74%. Dengan permasalahan diatas peneliti bermaksud untuk membuat sebuah sistem untuk melakukan analisis sentimen dengan Algoritme klasifikasi Naive Bayes. Dataset yang digunakan berupa teks kicauan (Tweet) yang bersumber pada media sosial Twitter dengan menggunakan Teknik Crawling. Proses pengumpulan data yang dilakukan menggunakan pustaka Tweepy dengan kata kunci “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah” yang dilakukan di rentan waktu 10 Mei 2022 hingga 1 Juni 2022. Selama proses pengumpulan data, peneliti mengambil attribute id, nama user, tanggal dan isi text. Setelah dilakukan pengumpulan data, tweet tersebut dilakukan labeling data agar memberikan label pada sentiment, kemudian tahap pre-processing dengan dilakukannya penghilangan simbol dan tanda baca yang tidak diperlukan, normalisasi data dilakukan dengan konversi kata yang semula tergolong ke dalam tidak baku menjadi kata baku, menghilangkan angka, dan penghapusan stopword. Kemudian masuk di tahap prediksi sentimen menggunakan metode Naive Bayes. Berdasarkan uraian tersebut penulis mengangkat judul topik penerapan algoritme Naive Bayes untuk analisis pandangan masyarakat terhadap pemindahan ibu kota indonesia.

Penelitian ini dilakukan dengan tujuan untuk melakukan pengklasifikasian tweet berbahasa indonesia terhadap pemindahan ibu kota ke kalimantan melalui media sosial Twitter dengan menggunakan Algoritme Naive Bayes dan untuk menguji tingkat akurasi dari Algoritme Naive Bayes dalam klasifikasi berbasis text.

2 Metode Penelitian

2.1 Data Penelitian

Twitter adalah salah satu media sosial terpopuler di dunia, dengan jumlah pengguna yang sangat banyak. Twitter memungkinkan penggunanya untuk mengirim pesan singkat yang disebut "tweet" yang dapat dilihat oleh orang lain. Tweet tersebut dapat berisi berbagai informasi, seperti pendapat, fakta, atau berita terbaru. Twitter menyediakan sebuah API (Application Programming Interface) yang memungkinkan pengguna untuk mengakses data tweet dan menggunakannya dalam berbagai aplikasi. Dengan menggunakan Twitter API, pengguna dapat mengambil data tweet secara real-time atau mengambil data tweet tertentu berdasarkan kriteria tertentu, seperti keyword, pengguna, atau lokasi. Dengan adanya Twitter API, para peneliti atau developer dapat mengolah data tweet untuk berbagai tujuan, seperti analisis sentimen, pemodelan bahasa, atau penelitian lainnya.

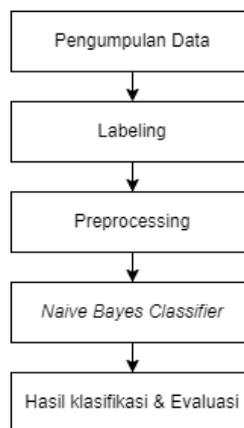
Pada penelitian ini data diperoleh menggunakan teknik crawling dengan bantuan pustaka Tweepy pada Python. Pengumpulan Data dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 dengan memasukan keyword “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah” sehingga diperoleh data dengan total 1.796.

Tabel 1. Sample Data Tweet

Full Name	Tweet	Date
yurikanamnam	@dutaSherliana Pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.	2022-05-11 02:36:42
Dewan Mahasiswa Justicia FH UGM	Wajah-wajah itu tidak datang dari ruang hampa. Melainkan kebijakan pemindahan ibu kota, yang pada akhirnya melatarbelakangi terukirnya ragam wajah dalam karya-karya yang kami sajikan. https://t.co/wlSNuiSaYI	2022-05-10 11:38:11
Dewan Mahasiswa Justicia FH UGM	yaitu mereka-mereka yang diuntungkan melalui kebijakan pemindahan ibu kota.\n\nDengan penyajian yang sederhana, barangkali potret wajah yang terpampang merupakan suatu yang mahal dan tak ternilai. Sebab, ialah kekuasaan menjadi harga yang harus dibayarkan.	2022-05-10 11:38:23
Margareth Saalam	Segala macam hoak diciptakan kelompok sakit hati terhadap Ibu Kota baru Negara. Tetap waspada. IKN Untuk Negeri https://t.co/0IE314G5IO	2022-05-23 02:05:25
Ramdani	@OposisiCerdas Yang jadi masalah Luhut mau bangun ibu kota baru....danannya dr hutang...yang bayar rakyat...dia dapat komisi dr investor....kan sialan dangkalan.....	2022-05-23 04:28:17
Anwar jahid Pena	Rakyat ga butuh Ibu kota baru..	2022-05-22 14:32:24

2.2 Penerapan Metode

Dalam membangun kerangka kerja pemeriksaan perasaan Twitter yang menggunakan teknik *Naïve Bayes*, diperlukan penggambaran keseluruhan yang menunjukkan perkembangan eksplorasi yang sedang diselesaikan sehingga siklus eksplorasi berjalan seperti yang diharapkan. Proses kerja ini membahas tahapan yang mendasari hingga batas terjauh dari kerangka kerja yang berjalan, yang terdapat pada Gambar 1 berikut:



Gambar. 1. Tahapan Metode Penelitian

Berdasarkan Gambar 1, Setidaknya ada 5 tahapan dalam penelitian ini dimulai dengan tahap pengumpulan data dengan memanfaatkan proses *crawling* data Twitter menjadi dataset mentah, kemudian tweet yang sudah berbentuk file .csv dilakukan proses *import* ke dalam program lalu kemudian dimasukkan kedalam database untuk dilakukan pelabelan kelas sentimen dan *preprocessing*. Pada tahap *preprocessing* dilakukan pembersihan, pemilihan dan pemilihan kata yang layak untuk diproses ke dalam tahap selanjutnya, Pada tahap *Naïve Bayes Classifier* dilakukan tahap pembagian data dengan ratio (90:10) untuk data latih dan data uji. Setelah itu dilakukan proses perhitungan *Document Frequency* (DF). Setelah perhitungan DF, kemudian dilakukan proses klasifikasi yang dilanjutkan dengan tahap selanjutnya yaitu menampilkan proses hasil dari klasifikasi dan evaluasi berupa hasil validasi akurasi atau perhitungan untuk menentukan nilai akurasi dan akurasi, dan hasil penerapan metode divisualisasikan dalam bentuk persentase.

2.3 Pengumpulan Data

Penelitian ini untuk mengumpulkan data yang digunakan dengan memanfaatkan teknik *crawling* Twitter. Teknik tersebut dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 dengan menggunakan *library* yang telah disediakan yakni Tweepy dan Twitter API. Twitter API dibutuhkan untuk menjalankan Tweepy sehingga data tweet dapat diakses melalui akun developer Twitter. *Crawling* data dilakukan berdasarkan keyword “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah”. Setelah itu dataset tweet yang masih mentah disimpan dalam bentuk csv file lebih lanjut.

2.4 Labeling Data

Proses pelabelan kelas sentimen pada setiap tweet dikategorikan dengan melihat karakter, pesan atau informasi yang dimuat pada sebuah tweet. Sistem penamaan (*Labeling*) harus dimungkinkan secara fisik oleh pencipta dengan penandaan kelas langsung mengingat subjektivitas kalimat secara individual pada tweet tanpa melalui sistem *preprocessing* terlebih dahulu. Penandaan perasaan dibedakan menjadi dua kelas, yaitu baik yang spesifik dan pesimistis. Kelas positif ditujukan untuk tweet yang berisi perasaan puas, ceria, dan hebat dengan hijrahnya Ibu Kota Negara ke Kalimantan Timur, sedangkan kelas negatif ditujukan pada tweet yang cenderung kecewa, tidak senang serta tidak puas dengan adanya pemindahan Ibu Kota Negara ke Kalimantan Timur. Setelah dilakukan proses pelabelan maka data yang akan disimpan di database akan dilakukan penyaringan sehingga akan menyisakan tweet yang akan dilakukan proses selanjutnya yaitu *preprocessing*.

2.5 Preprocessing

Informasi tweet yang didapat dari sistem *crawling* tidak dapat dimanfaatkan karena sebenarnya memiliki pesan atau struktur kalimat yang sporadis atau tidak terstruktur karena memiliki banyak sekali keributan. Oleh karena itu, pada tahap prapemrosesan ini, penting untuk mengubah jenis informasi dari informasi tidak terstruktur menjadi informasi terorganisir, bertekad untuk membuat tweet bersih dengan tujuan agar mereka dapat ditangani untuk mengeluarkan data dari laporan teks dan bekerja dengan fase siklus berikutnya. Tahapan *preprocessing* dalam penelitian ini meliputi:

- a. *Case folding*
Di dalam tahapan ini karakter-karakter yang mengandung huruf besar akan dikonversi menjadi huruf kecil (*lowercase*) seperti kata “Indonesia” diubah menjadi “indonesia”, dan lain sebagainya.
- b. *Cleansing*
Proses *cleansing* digunakan untuk menghilangkan atau menghapus gambar-gambar umum Twitter, misalnya username, @ (pemberitahuan), dan karakter yang berbeda seakan huruf utama (a-z) tertinggal. Sistem pemurnian ini memiliki beberapa fase, khususnya: menghilangkan karakter selain huruf (a-z), teks dengan 1 orang, URL, dan spasi berlebihan atau teks kosong sebagai spasi.
- c. *Mengganti Slang word*
Proses merubah *slang word* setiap kata kekinian atau yang tidak dianggap baku akan dikonversi menjadi bentuk bakunya, seperti kata “dri” menjadi “dari” dan lain sebagainya. Proses ini membutuhkan kamus *slang word* yang telah tersedia dalam basis data (database).
- d. *Menghapus Stopword*
Proses *stopword* dilakukan dengan menggunakan pustaka dari Nltk.corpus. Pada tahap ini, kata-kata tidak penting atau kata-kata umum tidak berdampak pada peristiwa kata-kata yang berkelanjutan. Pembatalan kata menggunakan *stopword* bahasa Indonesia yang telah disusun dan ditambahkan secara fisik mengingat masih ada beberapa kata yang belum diingat untuk *stopword* tersebut. Metode yang digunakan untuk menghapus kata-kata menyerupai model yang menyertainya, di mana kata-kata “ini”, dan “sekali lagi” ada di daftar kata berhenti, dan kata itu dihapus.

e. *Stemming*

Tahapan ini dilakukan dengan memanfaatkan pustaka dari Sastrawi, pada setiap kata yang memiliki imbuhan (berimbuhan) akan dikonversikan menjadi kata dasar seperti “makan” dihapus imbuhan “me-” sehingga menjadi kata “makan” dan lain sebagainya.

2.6 Naïve Bayes Classifier

Dari proses yang telah dilakukan sebelumnya telah didapatkan data dari hasil proses *preprocessing* yang kemudian akan diproses ke dalam klasifikasi, Hasil klasifikasi sentimen yang di dapatkan dengan menerapkan metode *Naïve Bayes Classifier*, terdapat 3 tahapan utama, yaitu pembagian dataset training dan testing, perhitungan DF (*Document Frequency*) dan klasifikasi.

a. *Pembagian Data*

Pada tahap ini, setelah dilakukannya *preprocessing*. Dataset dibagi dalam proses pembagian data ini dengan rasio 90 untuk data *training* berbanding 10 untuk data *testing*.

b. *Perhitungan TF-IDF*

Tahap ini adalah cara paling umum untuk menurunkan bobot istilah (kata-kata) ke catatan atau tweet untuk mengetahui relevansinya dalam korpus (berbagai laporan). Tahap pembobotan diingat untuk siklus vektorisasi yang diharapkan dapat mengubah teks dalam tweet menjadi angka numerik. Perhitungan TF-IDF menggunakan persamaan (1).

$$TF-IDF(t) = TF(t) * IDF(t) \tag{1}$$

yang mana:

- $TF(t) = (\text{jumlah kemunculan term } t \text{ dalam dokumen}) / (\text{jumlah total term dalam dokumen})$
- $IDF(t) = \log(N/n)$, dimana N adalah jumlah total dokumen dan n adalah jumlah dokumen yang mengandung term t

Dengan menggunakan rumus ini, nilai TF-IDF dari setiap term dapat dihitung, yang menunjukkan tingkat kepentingan dari setiap term terhadap dokumen. Semakin tinggi nilai TF-IDF suatu term, maka semakin penting pula term tersebut bagi dokumen tersebut.

c. *Klasifikasi Naïve Bayes Classifier*

Tahap karakterisasi dengan memanfaatkan *Naïve Bayes* merupakan tahap persiapan laporan yang telah diproses sebelumnya dan diberi tanda untuk mendapatkan hasil opini. Sistem pengelompokan ini membutuhkan informasi penandaan manual mulai dari penyiapan informasi hingga pembuatan model. Ada dua tahap dalam siklus karakterisasi, yaitu

1. Model dibangun berdasarkan analisa pada data *training*
2. Kemudian melakukan klasifikasi dari hasil model yang didapatkan akan dimanfaatkan kembali untuk dilakukannya klasifikasi terhadap label yang belum diketahui

Pada Algoritme *Naïve Bayes* setiap dokumen dipresentasikan dengan memasukkan “ $a_1, a_2, a_3, \dots, a_n$ ” di mana a_1 adalah kata pertama dan berikutnya sampai, sedangkan V yaitu label kategori, Selanjutnya yaitu mencari nilai tertinggi dari kategori teks yang diujikan (V_{MAP}) persamaan V_{MAP} yaitu pada rumus (2) sebagai berikut:

$$V_{MAP} = \underset{v_i \in V}{\text{argmax}} P(v_j) \cap_i P(a_i | v_j) \tag{2}$$

Nilai $P(v_j)$ dihitung pada saat data latih, dengan rumus (2) sebagai berikut:

$$P(v_j) = \frac{|Docs\ j|}{|training|} \tag{3}$$

$|Docs\ j|$ adalah jumlah dokumen yang memiliki kategori j pada dokumen latih dan $|training|$ adalah jumlah dokumen latih.

$$P(a_i|v_j) = \frac{|n_i + 1|}{|n + \text{kosakata}|} \quad (4)$$

Keterangan:

- n_i : Jumlah kemunculan kata a_i pada dokumen latih yang berkategori v_j
- N : jumlah seluruh kata pada dokumen latih yang berkategori v_j
- kosakataata : Jumlah kata unik pada seluruh dokumen latih

3 Hasil dan Pembahasan

Implementasi metode dalam penelitian ini dilakukan dengan empat (4) tahapan utama. Tahapan utama tersebut diproses secara berurutan, tahapan utama yang dimaksud antara lain: Tahap pengumpulan data, tahap labeling data, tahap preprocessing, tahap klasifikasi dengan Algoritme *Naïve Bayes Classifier*

3.1 Tahap Pengumpulan Data

Data yang dikumpulkan untuk penelitian ini menggunakan teknik *crawling* pada Twitter. Adapun teknik *crawling* dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 mengimplementasikan library yang telah tersedia yakni Tweepy dan Twitter API. *Crawling* data dilakukan berdasarkan keyword “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah”.. Setelah itu dataset tweet yang masih mentah disimpan dalam bentuk csv file.

Dataset yang dikumpulkan melalui proses *crawling* dengan informasi antara lain : tweet id, full_text (tweet), created_at (tanggal), tweet.user.name (username), yang akan disimpan ke dalam file excel (.csv) yang kemudian akan dimasukkan kedalam basis data (Database) secara *Import* Dataset didalam program berjumlah sebanyak 1.272 data tweet.

3.2 Tahap Labeling

Proses pelabelan (*labeling*) dilakukan dengan penamaan kelas langsung mengingat subjektivitas kalimat secara individual pada tweet tanpa melalui sistem pra-pemrosesan terlebih dahulu. Penandaan opini dipisahkan menjadi dua kelas, khususnya pasti dan negatif. Kelas positif ditujukan untuk tweet yang berisi perasaan terpenuhi, bahagia, dan hebat dengan migrasi Ibukota Negara ke Kalimantan Timur, sedangkan kelas negatif ditujukan pada tweet yang cenderung kecewa, tidak senang serta tidak puas dengan adanya pemindahan Ibu Kota Negara ke Kalimantan Timur. Berikut ini contoh dataset yang telah dilakukan proses labeling.

Tabel 2. Dataset Labeling

<i>Tweet</i>	<i>Label</i>
pindah kota perata bangun ekonomi indonesia infrastruktur bukti nyata	Positif
heboh tolak pindah ibukota	Negatif

Dari 1.272 dataset yang sudah dilakukan proses pelabelan (*labeling*), sebanyak 272 yang akan dilakukan tahap *Preprocessing* dan tahap *Classification*. Proses ini bertujuan untuk menyeimbangkan total tweet yang negatif dan positif agar mencapai nilai akurasi maksimal pada tahap *Classification* Menggunakan Algoritme *Naïve Bayes*.

3.3 Tahap Preprocessing

Setelah dataset dilakukan proses pelabelan dan proses maka selanjutnya dataset diproses dengan proses seperti *case folding*, menghapus teks dengan 1 karakter, menghapus URL, mengganti *slang word*, menghapus *stop word*, dan *stemming*.

a. Case Folding

Pada Tabel 3 merupakan proses untuk mengubah isi teks dokumen menjadi huruf kecil (*lower case*), melalui proses *case folding*, sehingga isi teks menjadi huruf kecil semua. Berikut contohnya:

Tabel 3. Case Folding

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
b'Rakyat...??? Masyarakat kaliii..... Pajak yg ditentukan juga biasa Ada di daerah lain.... Yang mau pindah silahkan... Yg gak mau pindah silahkan.... Gak Ada elu juga gak papa Tapi pemindahan ibu kota sangat2 tepat.... Agar pemerintah fokus melayani	b'rakyat...??? masyarakat kaliii..... pajak yg ditentukan juga biasa ada di daerah lain.... yang mau pindah silahkan... yg gak mau pindah silahkan.... gak ada elu juga gak papa tapi pemindahan ibu kota sangat2 tepat.... agar pemerintah fokus melayani

b. *Cleansing*

Pada Tabel 4 merupakan proses untuk menghapus karakter selain a-z. Sehingga isi teks yang bukan huruf a-z akan dihapus, seperti angka, simbol, dan seterusnya. Berikut contohnya:

Tabel 4. Cleansing

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
b'rakyat...??? masyarakat kaliii..... pajak yg ditentukan juga biasa ada di daerah lain.... yang mau pindah silahkan... yg gak mau pindah silahkan.... gak ada elu juga gak papa tapi pemindahan ibu kota sangat2 tepat.... agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak yg ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yg gak mau pindah silahkan gak ada elu juga gak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani

c. Mengganti *Slang Word*

Pada Tabel 5 merupakan proses untuk mengganti kata yang tidak baku (slang) menjadi kata baku, seperti kata “yg” dan “gak” diubah menjadi “yang” dan “tidak”. Berikut contohnya:

Tabel 5. Slang Word

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak yg ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yg gak mau pindah silahkan gak ada elu juga gak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak yang ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yang tidak mau pindah silahkan tidak ada lu juga tidak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani

d. Menghapus *Stop Word*

Pada Tabel 6 menggambarkan terjadinya proses penghapusan kata yang dikategorikan sebagai kata tidak terlalu penting dengan mengacu kepada kamus *stop word*, seperti kata dan, untuk, bagi, dan seterusnya. Berikut contohnya:

Tabel 6. Stop Word

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak yang ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yang tidak mau pindah silahkan tidak ada lu juga tidak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak ditentukan daerah pindah silahkan pindah silahkan lu papa pemindahan kota pemerintah fokus melayani

e. *Stemming*

Pada Tabel 7 menggambarkan penghapusan imbuhan dari tweet untuk menghasilkan cukup kata dasarnya saja. Berikut contohnya:

Tabel 7. Stemming

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak ditentukan daerah pindah silahkan pindah silahkan lu papa pemindahan kota pemerintah fokus melayani	rakyat masyarakat kaliii pajak tentu daerah pindah silah pindah silah lu papa pindah kota perintah fokus layan

3.4 Tahap Klasifikasi Naïve Bayes Classifier

a. Tahap Pembagian Data

Setelah proses *preprocessing* dilanjutkan dengan pemisahan data menjadi data latih dan data uji, maka proses pemisahan data dilakukan secara manual di dalam program. Untuk proses pemisahan data dengan perbandingan rasio 90:10, 90% untuk data latih dan 10% untuk data uji.

b. Perhitungan *Document Frequency*

Tahap perhitungan untuk mengubah teks dalam tweet menjadi angka yang dapat ditangani dan dicari desainnya selama proses pemesanan arsip. Langkah awal pembobotan adalah memecah kalimat menjadi berbagai kata (token), kemudian pada saat itu mengambil kata khusus untuk melakukan pembobotan. Seperti pada contoh Tabel 8 berikut:

Tabel 8. Proses Tokenisasi

Tweet	Token	Unik Token
pindah kota perata bangun ekonomi indonesia infrastruktur bukti nyata	pindah	pindah
	kota	kota
	perata	perata
	bangun	bangun
	ekonomi	ekonomi
	indonesia	indonesia
	infrastruktur	infrastruktur
	bukti	bukti
	nyata	nyata
	heboh tolak pindah ibukota	heboh
tolak		tolak
pindah		-
ibukota		ibukota
Total	13	12

Setelah mendapatkan kata yang unik dari setiap token, dilanjutkan dengan perhitungan TF. Setelah itu dilanjutkan dengan proses DF untuk kemudian diolah menjadi probabilitas yang akan digunakan untuk perhitungan prediksi sentimen. Proses perhitungan seperti pada Tabel 9 berikut:

Tabel 9. Perhitungan DF

Token	TF		DF
	d1	d2	
pindah	1	1	2
kota	1	0	1
perata	1	0	1
bangun	1	0	1
ekonomi	1	0	1
indonesia	1	0	1
infrastruktur	1	0	1
bukti	1	0	1
nyata	1	0	1
heboh	0	1	1
tolak	0	1	1
ibukota	0	1	1

c. Tahap *Naïve Bayes Classifier*

Dalam tahapan ini dilakukan proses *Naïve Bayes Classifier*, langkah pertama yaitu membuat model klasifikasinya dengan menghitung jumlah probabilitas kata untuk kategori positif dan negatif dengan dataset training sebanyak 244 dari 1272 data. Proses perhitungan jumlah probabilitas kata dengan membagi total kata dengan banyaknya kata, dan didapatkan hasil perhitungan sebagai berikut.

asd

Tabel 10. Perhitungan Probabilitas

Contoh Token	Probabilitas
pindah	0.1277139208173691
kota	0.3864406779661017
perata	0.023728813559322035
bangun	0.019157088122605363
ekonomi	0.03389830508474576
indonesia	0.14915254237288136
infrastruktur	0.008939974457215836
bukti	0.006385696040868455
nyata	0.006385696040868455
heboh	0.007662835249042145
tolak	0.010217113665389528
ibukota	0.05874840357598978

Setelah didapatkan nilai probabilitas pada setiap tokennya, langkah berikutnya yakni mengalikan keseluruhan nilai probabilitas masing-masing kategori untuk didapatkan batas nilai akhir dari setiap kategori nya, dengan hasil sebagai berikut:

$$P_{\text{kategori}} = \text{nilai probabilitas kategori} * (\text{total data kategori} / \text{total data training})$$

$$P_{\text{positif}} = 9,740540823468385e^{-18} * (120/244) = 7,295815054e^{-8}$$

$$P_{\text{negatif}} = 2,5475699800127307e^{-15} * (124/244) = 3,960415512e^{-7}$$

Setelah dilakukan perhitungan probabilitas maka didapatkan batas nilai akhir dari setiap kategori seperti pada hasil diatas. Apabila nilai probabilitas nya melebihi dari batas akhir nilai positif maka kata atau kalimat tersebut dinyatakan sebagai kata atau kalimat negatif.

4 Kesimpulan

Hasil dari penelitian ini dengan menyelesaikan tahapan perancangan, pembuatan sistem dan pengujian aplikasi, disertai dengan dataset juga algoritme yang diusulkan. Berdasarkan 1.272 tweet, arah pandangan (sentimen) masyarakat indonesia terhadap pemindahan ibu kota indonesia cenderung kearah sentimen positif sebesar 86.95 % pada periode Mei – Juni 2022. Tahapan utama dalam penelitian ini adalah *Crawling, Labeling, Preprocessing* dan *Classification* menggunakan *Naïve Bayes Classifier*. Tahap *Preprocessing* yang baik menjadi penentu dalam terbentuknya hasil yang optimal untuk tahap selanjutnya. Penggunaan algoritme *Naïve Bayes Classifier* dalam melakukan analisis sentimen dapat berjalan dengan baik, dengan nilai pengujian dan evaluasi tertinggi yang diperoleh sebesar: akurasi 93%, presisi 87%, dan recall 100%.

Referensi

- [1] A. M. Hidayat, dan M. Syafrullah, "Algoritma Naïve Bayes dalam Analisis Sentimen untuk Klasifikasi Pada Layanan Internet PT. XYZ," *Jurnal Telematika MKOM*, vol. 9, no. 2, hal. 91–95, 2017.
- [2] Bappenas, Dampak Ekonomi dan Skema Pembiayaan Pemindahan Ibu Kota Negara, 2019.
- [3] F. E. Purwiantono, dan A. Aditya, "Klasifikasi Sentimen Sara, Hoaks dan Radikal pada Postingan Media Sosial Menggunakan Algoritma Naïve Bayes Multinomial Text," *Jurnal Teknopak*, vol. 14, no. 2, hal. 68–73, 2020.
- [4] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2021.
- [5] S. Afrizal, "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit," *JURNAL INFORMATIK*, 2019.
- [6] Afrizal, S., Irmanda, H.N. and Falih, N, "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit," hal. 157–168, 2019.

- [7] P. Antinasari, R. S. Perdana, dan M. A. Fauzi, “Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku”, hal. 1733–1741, 2017.
- [8] P. Arsi dan R. Waluyo, “Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, hal. 147–156, 2021.
- [9] A. R. Isnain, N. S. Marga, dan D. Alita, “Sentiment Analysis of Government Policy On Corona Case Using Naive Bayes Algorithm,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, hal. 55, 2021.
- [10] M. Priandi dan Painem, “Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,” dalam *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, 2021.
- [11] A. V. Sudiantoro, dan E. Zuliarso, “Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier”, hal. 69–73, 2018.